



HARVARD Kennedy School

**MOSSAVAR-RAHMANI CENTER**  
for Business and Government

# **Unleashing the potential of randomised controlled trials in Australian governments**

**Phil Ames**  
**James Wilson**

April 2016

## **M-RCBG Associate Working Paper Series | No. 55**

The views expressed in the M-RCBG Associate Working Paper Series are those of the author(s) and do not necessarily reflect those of the Mossavar-Rahmani Center for Business & Government or of Harvard University. The papers in this series have not undergone formal review and approval; they are presented to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

# Contents

Background and Report Structure.....	2
One Page Summary of Strategy for RCTs in Social Policy .....	4
<b>1. Context.....</b>	<b>5</b>
1.1. Introduction to RCTs.....	5
1.2. Experience from around the world .....	9
1.3. Australian Context.....	12
<b>2. Start RCTs Now .....</b>	<b>15</b>
2.1. Specific recommended interventions for early RCTs .....	16
2.2. Criteria for which RCTs to pursue.....	18
2.2.1. Good data– Can you accurately measure the outcome using existing data? .....	18
2.2.2. Suitable populations – Can you, in a fair way, randomise a reasonably large population? .....	19
2.2.3. Strategic interventions – Are interventions standardised, easily varied, and impactful?... ..	20
<b>3. Build RCT Capability .....</b>	<b>22</b>
3.1. Culture & Leadership .....	23
3.1.1. Institutional measures.....	24
3.1.2. Communication strategies .....	25
3.2. Expertise.....	27
3.2.1. Centre of Excellence.....	27
3.2.2. Partnerships.....	30
3.3. Transparency.....	32
3.3.1. Public Registry .....	32
3.3.2. Link Administrative data .....	33
<b>Conclusion.....</b>	<b>35</b>
<b>4. Appendix .....</b>	<b>36</b>
4.1. Survey of Australian Parliamentary Views of RCTs .....	36
4.2. Table of minimum sample sizes .....	37
4.3. Australian RCTs in Social Policy.....	38
4.4. Successful International Social Policy RCT Examples .....	42
4.5. Research Appendices .....	43

# Background and Report Structure

The authors (Phil Ames and James Wilson, hereafter ‘we’), have completed this report as our Policy Analysis Exercise (PAE) in our Master in Public Policy degrees at the Harvard Kennedy School.

Our lead advisor is the Hon Dr Andrew Leigh MP, a current Member of the Australian Federal Parliament, the Federal Member for Fraser, the Shadow Assistant Treasurer and Shadow Minister for Competition for the Australian Labor Party, with a Ph.D. and M.P.A. from Harvard Kennedy School. Leigh is a Fellow of the Australian Academy of Social Sciences and has authored numerous books including *Disconnected* (2010), *Battlers and Billionaires* (2013), *The Economics of Just About Everything* (2014) and *The Luck of Politics* (2015).

Leigh is passionate about improving the effectiveness of governments in Australia, and to that end wants to see randomised controlled trials (RCTs) used more often in social policy. As many as 13 years ago, while completing his Ph.D. at Harvard Kennedy School, Leigh wrote a paper on this topic, concluding with the call:

*“Australian policymakers should summon up the vigour of their predecessors, and conduct randomised trials on a variety of current and proposed policies — providing evidence on what works, and what does not.”<sup>1</sup>*

Now, as one of those policymakers summoning vigour, Leigh put to us the challenge of designing a practical, pragmatic strategy for increasing the use of RCTs in Australian governments. We organised our work around this central question:

***What should be the pragmatic strategy for the Australian Federal Government to transition to greater use of RCTs in social policy?***

Given the potential for RCTs to improve social policy outcomes not just for the Australian Federal Government, but for governments and agencies across Australia and abroad, we have attempted to make our findings and recommendations of relevance to policymakers and senior public servants across contexts.

While the primary audience for this report is the Australian Federal Government, we intend to distribute our findings and recommendations widely both within Australia and around the world in the hope of supporting a transition to greater use of RCTs in social policy.

Our thanks to our advisors from Harvard, Max Bazerman, Mitch Weiss and John Haigh.

---

<sup>1</sup> Leigh, “Randomised Policy Trials.”

This document is divided into three sections:

**1. One Page Summary of Strategy for more RCTs in Social Policy**

This summarises our findings and recommendations for the Australian Federal Government.

**2. Details and Analysis**

This section outlines the full rationale and detail of our findings and recommendations, following the summary's structure. While globally relevant, the focus of this section remains on the Australian Federal Government.

**3. Appendix**

The Appendix contains the more voluminous content underlying our findings and recommendations. It includes:

- Summary of findings from our Survey of Australian Parliamentarians' Views on RCTs in Social Policy
- Table of minimum sample sizes for RCTs
- Tables and examples of RCTs around the world and in Australia
- Research appendices (methodology, interview list and references)

---

*"Our failure to assess the effectiveness of government's spending carries more than a financial cost – it has massive human cost as well. When a program designed to boost employment fails to perform, parents are unable to put food on their families' tables. When an education initiative doesn't live up to its promise, neither can the young children enrolled in it."*<sup>2</sup>

---

---

<sup>2</sup> Nussle and Orszag et al., *Moneyball for Government*, 5.

# One Page Summary of Strategy for RCTs in Social Policy

There is momentum around the world for running RCTs both to optimise the delivery of existing services and to test the effectiveness of social policy. With only a limited history running such RCTs, there is now an opportunity for Australian governments to follow the path increasingly set by both the UK and USA and significantly expand the use of RCTs in social policy. In the last three years, momentum has been growing for using RCTs in government in Australia, particularly in the Federal and NSW governments.

To capitalise on that momentum and build a culture of testing using RCTs, we have developed the recommendations below for Ministers and Senior Public Servants. They are organised around starting early and learning from experience, and then building the capability of the public service to run more RCTs in the longer term. Below are our recommendations, each of which is explained in greater detail throughout this document.

We recommend:	Ministers	Senior Public Servants
Start RCTs Now		
Early RCTs	Start by investigating an RCT for an intervention in social policy. Start with interventions that are likely to be successful, and start in areas likely to provide value for money for government. We recommend starting with one of: <ul style="list-style-type: none"><li>Cognitive behavioural therapy to reduce crime</li><li>Meta-cognition and self-regulation programme to improve educational outcomes</li></ul>	Start by commissioning RCTs on small, uncontroversial, high impact/cost likelihood, service optimisation interventions. We recommend starting with one of: <ul style="list-style-type: none"><li>Moving the signature box to the top of forms involving disclosure to government</li><li>Increasing collection of owed payments to government using behaviourally-informed letters</li><li>Increasing payment of fines using behaviourally-informed SMS reminders</li></ul>
Apply criteria for which RCTs to pursue	1. Good data – Can you accurately measure the outcome of interest using existing data? 2. Suitable populations – Can you, in a fair way, randomise a reasonably large population? 3. Strategic interventions – Are interventions standardised, easily varied, and impactful?	
Build RCT capability		
Culture and Leadership	Demand evidence and RCT evaluation strategies from public servants and allocate funding based on evidence	Push available evidence to ministers and outline the most rigorous possible evaluation strategies (including RCTs) for new policies
	Request that each year every agency identify: <ul style="list-style-type: none"><li>Three policy areas with limited evidence of effectiveness</li><li>Actions to be taken to increase the evidence base</li><li>At least one RCT they intend to run</li><li>Results from the previous year’s RCT(s)</li></ul>	
Expertise	Establish BETA as a ‘centre of excellence’ for RCTs which collaborates with departments to provide RCT expertise for the Federal Government	
	Better engage the academic community and social sector	
Transparency	Authorise a public registry of policy RCTs	Link administrative data within departments, between departments and across jurisdictions

Table 1 - Summary of Recommendations

# 1. Context

This 'Context' section:

- Provides a definition of RCTs and describes where they fit in the policy-making process;
- Summarises the state of play from the UK, US and international development fields; and;
- Reviews the use of social policy RCTs in Australia so far.

For this section we attempted to establish a view on the empirical base of social policy RCTs in Australia and around the world. This has been challenging for several reasons. To begin with, not all social policy RCTs are published. Moreover, creating a definitive view on published RCTs is challenging because not all RCTs are on registries and some RCTs are on multiple registries. This makes attempts to reconcile numbers across public registries fraught. We have not sought to do this, but rather to analyse each repository for what it offers. As a consequence, the numbers we cite represent a best-endebours snapshot, rather than a collectively exhaustive, mutually exclusive view of all published, social policy RCTs. See the *Research Appendices* for a description of the databases we analysed.

## 1.1. Introduction to RCTs

### ***What is an RCT?***

A randomised control trial (RCT) is “a study that randomly assigns individuals or other units (such as schools or counties) to one group that is eligible to participate in a programme [or intervention], or to a ‘control group’ that is not.”<sup>3</sup> With large enough sample sizes, this approach allows both the treatment and the control group to be very similar in terms of observable and unobservable characteristics. Differences in outcomes between the treatment and control groups can then be causally inferred, as there are no other differences between the groups.

The figure below demonstrates how an RCT is conducted.<sup>4</sup> A population is split into two groups through a randomisation process, such as a lottery. One group receives the intervention (the “treatment” group), whilst the “control” group does not. Outcomes for both groups are then measured. When the randomisation is carried out correctly, the two groups are the same across observable and unobservable characteristics, allowing any difference in outcomes to be attributed to the intervention.

---

<sup>3</sup> “Rigorous Program Evaluations on a Budget: How Low-Cost Randomised Controlled Trials Are Possible in Many Areas of Social Policy,” March 2012, 1.

<sup>4</sup> Haynes et al., “Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials,” 4.

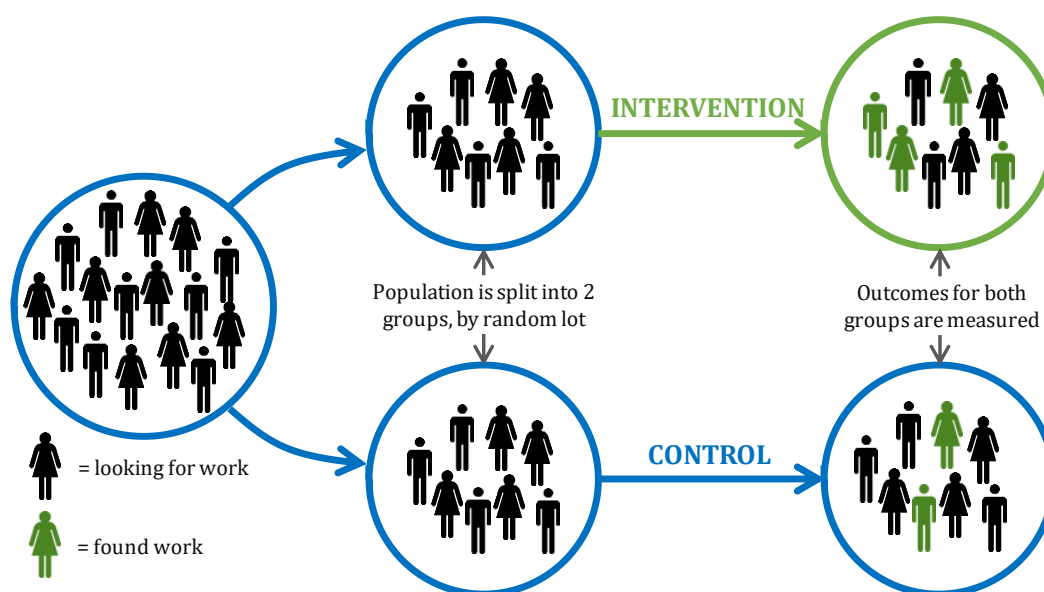


Figure 1 - How an RCT is conducted (from Test, Learn, Adapt by the Behavioural Insights Team)

The key advantage of RCTs over other evaluation methods is the randomisation process. In programmes where participants can self-enrol, the treatment and control groups may differ in unobservable ways, such as motivation levels. Other study designs without control groups make it very difficult to know if it was the intervention or another factor which led to a change in outcomes. As a consequence, the US National Academy of Science recommends that evidence of programme effectiveness generally “cannot be considered definitive” without a well-conducted RCT.<sup>5</sup>

There are further advantages to RCTs in the social policy realm. First, the prospect of definitive evidence promises to address contentious policy debates. For example, the Obama administration has used RCTs to evaluate teen pregnancy programmes.<sup>6</sup> RCTs allow many parties to put forward an approach to tackling a problem, test multiple approaches, and work out which one works.<sup>7</sup> This evidence-based policy approach is in contrast to current “policy-based evidence” approaches, where evidence is used selectively to support a pre-determined policy program.<sup>8</sup>

Furthermore, governments are well-placed to run RCTs.<sup>9</sup> Governments typically have substantial resources, and they generally run programmes that impact large populations, allowing for sufficient sample sizes for a treatment and a control group. In addition, governments already collect administrative data from citizens, reducing the need to create bespoke evaluation data. And unlike other service providers who may not have the resources to scale their intervention, governments can trial programmes that they would also scale, increasing the importance of rigorous evaluation.

<sup>5</sup> “Rigorous Program Evaluations on a Budget: How Low-Cost Randomised Controlled Trials Are Possible in Many Areas of Social Policy,” March 2012, 1.

<sup>6</sup> Buck and McGee, “Why Government Needs More Randomised Controlled Trials: Refuting the Myths,” July 2015, 2.

<sup>7</sup> Mullainathan and Faye, “Field Experiments in Rural Finance: An Example from Tamil Nadu, India,” 5.

<sup>8</sup> Haskins and Margolis, *Show Me the Evidence*, 2015, 5.

<sup>9</sup> Glennerster, “So You Want to Do an RCT with a Government: Things You Should Know — Running Randomised Evaluations: A Practical Guide.”

In this paper, we will be referring to two types of RCTs:

- **Optimisation RCTs** seek to improve the effectiveness of existing policy activities. These often relate to communications – testing different emails or letters for their relative effectiveness.
- **Policy RCTs** seek to test the effectiveness of new or existing policies, programs or related interventions. These range from trials of health insurance to trials of cognitive behavioural therapies to reduce recidivism.

### ***When is the use of RCTs appropriate?***

Some areas of government policy lend themselves to rigorous evaluation through the use of RCTs. Typically, these areas will be marked by the following features:

- *There is a defined and observable outcome measure:* A sex education policy in schools may have as an outcome measure reducing teenage pregnancy. This is a defined and observable goal – it is possible for researchers to track pregnancy rates amongst the cohort of students under treatment and control arms.
- *There are large populations under treatment:* In order to statistically distinguish between the groups, there needs to be many participants. To continue the example, there are thousands of schools with scores of students that receive sex education.
- *There is a standardised intervention:* For RCTs to deliver meaningful evidence as to what was done to improve the outcome, all participants in the treatment group must receive the same, standard intervention. For example, delivering the sex education curriculum through a computer programme would be a more standardised approach than providing a range of possible curricula to be chosen and delivered by the schools.

However, for some policy areas, the use of RCTs would be infeasible and inappropriate. In these instances, other evaluation methods should be used, which are not the subject of this report. Two factors that frequently preclude the use of RCTs are:

- *It is not possible to have a control group:* It may either be infeasible or unethical to separate off a control group. For example, it would be infeasible for the government to lower or raise interest rates in one randomly selected part of the country but not another.
- *The sample sizes are too small:* For example, a policy targeting foster children in rural NSW may not have enough participants to randomise.

While these constraints and limitations are important to understand, we overall agree with the many economists who contend that RCTs are “radically under-applied”<sup>10</sup> and that the “biggest problem with RCTs is that they are not used nearly often enough.”<sup>11</sup>

---

10 Cowen, “How Good Are RCTs in Economics? - Marginal Revolution.”

11 “In Praise of Human Guinea Pigs.”



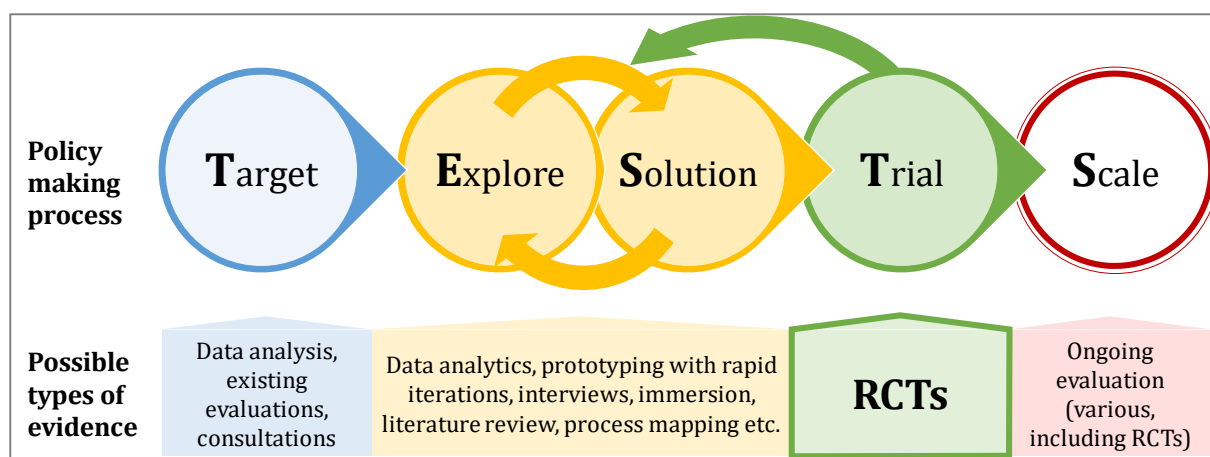


Figure 2 - RCTs should complement the policy making process<sup>12</sup>

Furthermore, as shown in Figure 2 above, RCTs should complement existing evidence-based policy-making processes, not replace them. Nor should they replace other forms of evidence, which in many parts of the policy-making process are more appropriate than RCTs.

Prominent critics argue that RCTs stifle innovation throughout the policy-making process. For example, Ricardo Hausmann contends that RCTs:

*“make us think about interventions, policies, and organisations in the wrong way. As opposed to the two or three designs that get tested slowly by RCTs (like putting tablets or flipcharts in schools), most social interventions have millions of design possibilities and outcomes depend on complex combinations between them. [...] Only by creating organisations that learn how to learn..., can we accelerate progress.”<sup>13</sup>*

Michael Barber, former advisor to UK Prime Minister Tony Blair, has similarly argued that “too often [evidence] is used to justify incrementalism or delay” and “the research is never complete and cannot tell you whether an innovation will succeed or not.”<sup>14</sup>

We agree with these concerns for whether RCTs promote the development of a learning culture. Ideally, RCTs should follow an iterative approach with design-led intervention prototypes having ample opportunity to explore different intervention mechanisms and combinations. Many RCTs should require “some level of adaptation to ensure that programmes ‘fit’ local conditions.”<sup>15</sup> Yet the risk remains that RCT evidence can instead lead to interventions being replicated without the ongoing testing or innovation central to developing a learning culture.

However, a recent report by the former Secretary of the Department of Prime Minister and Cabinet in Australia, Peter Shergold, outlines how RCTs can facilitate a learning culture. In recommending the use of RCTs as a component of more innovative policy-making, Shergold argues that “new policy proposals should include a trial or demonstration stage, allowing new approaches to be developed fast and evaluated early.”<sup>16</sup> This view of RCTs as enabling government to be nimble in their evaluation of policy proposals also stands as an important counter-balance to Barber’s views noted earlier.

As outlined in detail below, RCTs are currently “being applied to problems that once seemed off-limits (such as policing and education). They have got much bigger... [and] are even spreading to

12 Adapted from the UK Behavioural Insights Team.

13 Hausmann, “The Problem With Evidence-Based Policies.”

14 Barber, *How to Run a Government*, 2015, 7–8.

15 Cherney and Head, “Evidence-Based Policy and Practice Key Challenges for Improvement,” 521.

16 Shergold, “Learning from Failure,” 82.

rich countries.”<sup>17</sup> We think this growing use and application of RCTs around the world supports Shergold’s view that incorporating trials into the heart of policy-making is both viable and valuable.

## 1.2. Experience from around the world

We have developed *Table 2* below, which estimates the distribution of social policy RCTs across countries and over time. It is based on data from the Campbell Collaboration, “an international research network that produces systematic reviews of the effects of social interventions in Crime & Justice, Education, International Development, and Social Welfare.”<sup>18</sup> These systematic reviews collate high-quality randomised policy trials from around the world from many sources and databases. Note that this source exclusively identifies high-quality RCTs that are relevant to prescribed topics of interest. Accordingly, as mentioned above, these numbers will differ from other counts of RCTs in this report.

<i>Country and Year</i>	<i>Pre-’85</i>	<i>’85-94</i>	<i>’95-04</i>	<i>’05-14</i>	<i>Total</i>	
<b>Top Countries</b>						
<i>US</i>	81 (88%)	143 (87%)	227 (68%)	78 (29%)	529	62%
<i>Canada</i>	0 (0%)	8 (5%)	24 (7%)	5 (2%)	37	4%
<i>UK</i>	3 (3%)	0 (0%)	11 (3%)	20 (7%)	34	4%
<b><i>Australia</i></b>	<b>2 (2%)</b>	<b>3 (2%)</b>	<b>15 (5%)</b>	<b>13 (5%)</b>	<b>33</b>	<b>4%</b>
<i>India</i>	3 (3%)	2 (1%)	2 (1%)	13 (5%)	20	2%
<i>Kenya</i>	0 (0%)	0 (0%)	7 (2%)	11 (4%)	18	2%
<i>Mexico</i>	0 (0%)	0 (0%)	9 (3%)	9 (3%)	18	2%
57 other countries	3 (3%)	9 (5%)	37 (11%)	118 (44%)	167	20%

*Table 2 - Summary of cited RCTs by Campbell Collaboration Systematic Reviews across Social Policy Areas*<sup>19</sup>

This table reveals the dominance of North America in producing policy RCTs. The UK is in third place on policy RCTs, and also has a recent history of leading practices in optimisation RCTs. Australia holds fourth spot on this table, a spot it held when a similar analysis was performed in 2003.<sup>20</sup> This table also shows the emergence of three developing countries in the last decade: India, Kenya and Mexico. This reflects the growth of RCTs in the world of international development. We discuss each of these regions below.

### ***The US***

To better understand the US’ policy RCT landscape, we have analysed the American Economic Social Science Registry, setup in 2012.<sup>21</sup> It has 564 registered RCTs, of which approximately 150 are US-based with the remainder occurring in other jurisdictions.

As *Figure 3* shows, nearly three-quarters of the registered RCTs in the US are commissioned by academic institutions; only one was registered directly by a government department. This demonstrates that there is much expertise and enthusiasm for RCTs within US academia.

<sup>17</sup> “Randomised Control Trials”

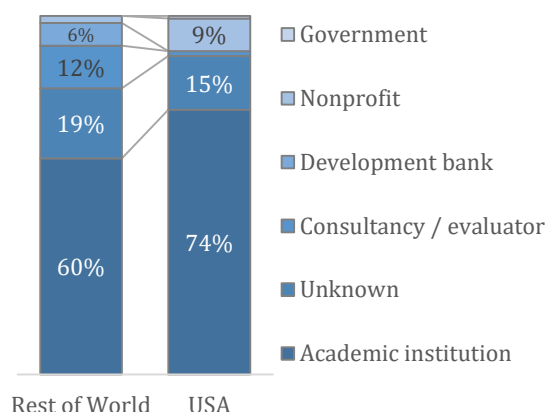
<sup>18</sup> The Campbell Collaboration, “Campbell Collaboration Systematic Reviews: Policies and Guidelines,” March 2, 2015.

<sup>19</sup> The Campbell Collaboration, “Campbell Collaboration Systematic Reviews: Policies and Guidelines,” March 2, 2015. This table is based on an extract from all systematic reviews from The Campbell Collaboration at the beginning of 2016. The results directionally align with a similar analysis performed in 2003 by Leigh, in which the USA was estimated to have produced 93% of policy RCTs, followed by Canada, the UK then Australia. While the order has remained the same, we have noted a reduction in the global dominance of US RCTs from 93% to 29% between 2005 and 2014. For Leigh’s paper, see Leigh, Andrew. “Randomised Policy Trials.” *Agenda* 10, no. 4 (2003): 341–54.

<sup>20</sup> Leigh, “Randomised Policy Trials.”

<sup>21</sup> “AEA RCT Registry.”

### AEA RCT Registry: Commissioning Institutions



### AEA RCT Registry: Policy Areas

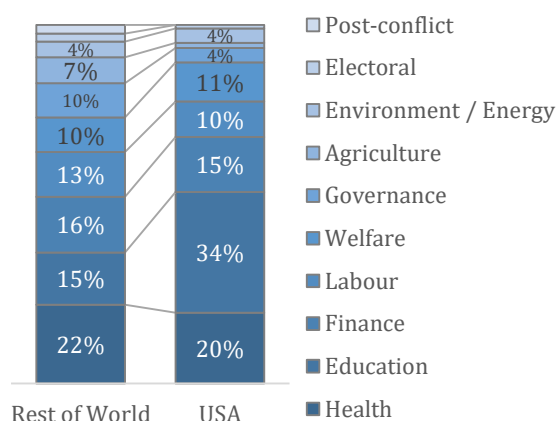


Figure 3 - Distribution of commissioning institution in AEA database  $n = 564$ <sup>22</sup>

In addition, *Figure 3* shows that in the US, over 50% of RCTs have been in either education or health. In the RCTs from the rest of the world, however, there is quite an even distribution across policy areas including finance, labour, welfare and governance.

The US Federal Government has proactively been supporting the growth of RCTs in recent years, including through the Social Innovation Fund, Investing in Innovation Fund (i3) and a range of pay-for-success social investments.<sup>23</sup> Most recently, the US created the Social and Behavioral Sciences Team (SBST) – launched publicly in September 2015.<sup>24</sup> This team has the same ethic of rigorous evaluation that the UK's Behavioural Insights Team (discussed below) pioneered. In its first year, the SBST focused on building credibility through low-cost, immediate and easily quantifiable interventions in government policies.<sup>25</sup> These have mostly taken the form of optimisation RCTs. These interventions fell into two categories: streamlining access to programmes and improving government efficiency. Both positive trial results and trials that resulted in no discernible impact are reported. The following are some examples of their interventions:<sup>26</sup>

- \$1.3m saving in the first month from an email campaign to increase enrolment in the Thrift Savings Plan, a workplace savings plan for federal employees
- 5.7 percentage point increase in college enrolment for low-income students through sending eight personalised text messages to prompt completion of pre-matriculation tasks
- No difference in payment rates for outstanding non-tax debt by redesigning a collection letter, but a 45% increase in the use of an online tool from shortening the URL in the letter

### The UK

The most well-known source of RCTs in government is the UK Government's 'nudge unit', also known as the Behavioural Insights Team (BIT). BIT started in the Prime Minister's office in 2010 as the world's first government institution dedicated to behavioural sciences and the regular use of RCTs. In 2014, BIT became a social purpose company, jointly owned by the Cabinet Office, Nesta

<sup>22</sup> Ibid.

<sup>23</sup> "Building the Evidence Base for What Works."

<sup>24</sup> "Executive Order -- Using Behavioral Science Insights to Better Serve the American People."

<sup>25</sup> "Social and Behavioral Sciences Team Annual Report," September 2015, 11.

<sup>26</sup> Ibid., 11–13.

and BIT employees.<sup>27</sup> Their approach is “highly empirical” – testing and trialling all ideas through a “test, learn, adapt” model.<sup>28</sup> BIT helped the UK government conduct over 165 RCTs in 2015, predominantly to optimise the delivery of existing services.

In addition to BIT, the UK has also established a network of What Works Centres. The most prolific of these, the Education Endowment Foundation (EEF) is an independent grant-making charity that has commissioned 119 RCTs on interventions to close the attainment gap in the UK’s education system.<sup>29</sup> To support the theory of change that better evidence is crucial, EEF funds three stages of tests:

- **Pilots** to refine an intervention that is at an early or exploratory stage of development;
- **Efficacy trials** to see whether an intervention can work under ideal or developer-led conditions; and;
- **Effectiveness trials** to test whether an intervention is effective under realistic conditions in a large number of schools.<sup>30</sup>

This three stage approach is a commitment to continual testing, learning and improvement – not a “one and done” approach. In total, the EEF has funded trials in “more than 4,000 UK schools and involving more than 600,000 children.” This means that the EEF has “conducted more large-scale trials than have been conducted across the entire education sector *ever*.”<sup>31</sup>

A further part of this commitment to learning is the transparency of their trial results. Of the 119 trials shown on EEF’s website, 45 are complete. Of those, 11 show statistically significant positive impacts. This transparency builds credibility in trial results and allows other jurisdictions to learn from EEF’s work.

### ***International Development***

Finally, in the international development arena, as shown in *Table 2*, RCTs have become much more common in the past decade. This has been spearheaded by the Abdul Latif Jameel Poverty Action Lab (JPAL), which has commenced or completed over 700 RCTs.<sup>32</sup> For example, an RCT conducted in India on whether giving a goat or cow to a family improved outcomes showed overwhelming success after five years.<sup>33</sup> Differential recruiting advertisements for teachers in Zambia led to teachers attending twice as many community meetings and undertaking 29% more home visits.<sup>34</sup>

---

27 “The Behavioural Insights Team, Update Report 2013-2015,” 5.

28 Haynes et al., “Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials.”

29 “About.”

30 “Evaluation Glossary.”

31 Halpern, *Inside the Nudge Unit*, 285.

32 “Measure for Measure.”

33 “In Praise of Human Guinea Pigs.”

34 Duflo and Karlan, “What Data Can Do to Fight Poverty.”

## 1.3. Australian Context

2016 is a time of opportunity for RCTs in Australian governments. While generally, for both optimisation RCTs and policy RCTs, Australian government departments do not systematically use randomised methods for evaluation, there is fresh momentum being driven by the proponents of behavioural science in government.

At a state level, the NSW Department of Premier and Cabinet's (DPC) Behavioural Insights Unit (BIU) was the first such unit in Australian government, having been launched in 2013. Federally, this was followed by the Department of Prime Minister and Cabinet's Behavioural Economics Team of Australia (BETA), launched in 2015 with 10 partner agencies and departments.<sup>35</sup> In March 2016, Victoria followed, announcing a new behavioural insights office in the Department of Premier and Cabinet.<sup>36</sup> While a headline role of these teams is to spread the application of behavioural science in government, an equally important priority is to run "major trials".<sup>37</sup> Their arrival in Australia is particularly welcome.

There is also momentum for RCTs beyond the behavioural science teams. As mentioned above, Shergold recently suggested in his recommendations for "Adaptive government" that the federal government should "Undertake controlled trials of government policy."<sup>38</sup> Further, our survey of 109 Australian Parliamentarians found that 73% are supportive of the use of RCTs in social policy, and 51% prioritised RCTs as a top three input to which politicians should pay attention (significantly up from only 23% in the UK).<sup>39</sup>

Australia clearly has momentum to build on, and there is an opportunity to take those RCTs that have already been conducted in Australia as a proof-of-concept from which to scale to wider implementation of RCTs in policy.

### ***Optimisation RCTs***

Australia is in the early days of using RCTs to optimise existing activities. Most agencies and departments in most governments do not currently run RCTs. The primary exception so far in Australia, is the NSW BIU. While BETA is yet to publicise its trials, in the last two years the NSW BIU has run at least seven RCTs optimising existing government activities, and there are likely more currently underway. The RCTs run by the BIU in NSW so far include working with:<sup>40</sup>

- Family and Community Services to send SMS messages to encourage tenants to "pay back their arrears quicker, sign up to an arrears payment plan or the Rental Deduction Scheme"
- The Office of State Revenue to make notices of taxes, fines and debts easier to understand and pay
- Allianz and the Department of Education and Communities to help "injured employees get back to work sooner"
- The Office of Environment and Heritage to "increase the uptake of the Home Power Savings Program"
- The Ministry of Health's Centre for Preventative Health to "increase the take up of workplace health screening and other referral services"
- Cancer Institute NSW to "test how to increase the take up of cervical screening"<sup>41</sup>

---

35 "Turnbull Government Creates New 'Behavioural Economics' Team of Advisers."

36 "Victoria Gives Behavioural Insights a Nudge, Cross-Agency Mandate."

37 "Understanding People, Better Outcomes - Behavioural Insights in NSW," 2.

38 Shergold, "Learning from Failure," 67.

39 See *Appendix Table 9* for full survey results.

40 "Understanding People, Better Outcomes - Behavioural Insights in NSW," 4.

41 "Spotlight on Health Results Behavioural Insights Short Report," 2.

- St Vincent's Hospital to reduce missed outpatient appointments<sup>42</sup>

While this momentum is promising, it pales in comparison to the UK BIT's 165 RCTs in the UK in 2015. Most states are not running RCTs, and in NSW and the Federal Government they are only being run in a limited number of contexts. There is great opportunity for expansion.

### ***Social Policy RCTs***

As outlined earlier in *Table 2*, we estimate that Australia remains the fourth highest producer of RCTs in social policy in the world. Combining research from the Campbell Collaboration with a broader search of academic databases, government reports and media, we have compiled a table of all major RCTs in social policy in Australia. The results are outlined in *Appendix 4.3 Australian RCTs in Social Policy* and summarised below in *Table 3*. Our research identified major RCTs in social policy in Australia, of which 41% were closely related to health or public health (excluding clinical trials).<sup>43</sup> There has been an increase in the number of social policy RCTs since the early 1990s, and a slight diversification in the policy areas covered.

<b><i>RCT Policy area</i></b>				
<i>Note: RCTs may be included in more than one policy area</i>				
	<b><i>Pre-'95</i></b>	<b><i>'96-'05</i></b>	<b><i>'06-15</i></b>	<b><i>Total</i></b>
Health or Public Health	5 (50%)	12 (41%)	13 (39%)	30 (41%)
Parenting / Family	1 (10%)	7 (24%)	5 (15%)	13 (18%)
Justice / Police	0 (0%)	4 (14%)	4 (12%)	8 (11%)
Education	2 (20%)	3 (10%)	3 (9%)	8 (11%)
Indigenous	1 (10%)	1 (3%)	3 (9%)	5 (7%)
Welfare	0 (0%)	2 (7%)	0 (0%)	2 (3%)
Other	1 (10%)	0 (0%)	5 (15%)	6 (8%)
<b>Total</b>	<b>10</b>	<b>29</b>	<b>33</b>	

*Table 3 – Estimated history of social policy RCTs in Australia by policy area and decade*

This table shows that RCTs in social policy in Australia are viable. They have been run in welfare, justice, policing, parenting and public health. Some were led by academics and others were run in partnership with government departments. However, the use of such RCTs as evidence for social policy remains the exception in Australia. As summarised by Deborah Cobb-Clark, Professor of Economics at the University of Sydney, “our current evaluation system generally produces poor-quality evaluations that in the end do not tell us very much.”<sup>44</sup>

One dynamic underlying the rarity of RCT use in policy in Australia is that the recent growth in interest in ‘evidence-based’ policy has not translated into interest in RCTs. *Figure 8* below illustrates that the explosion of interest in ‘evidence-based’ policy has not been followed by discussion of RCTs in Parliament nor in the media.

<sup>42</sup> Ibid.

<sup>43</sup> This table was built from a combination of academic databases, government websites, media searches and the Campbell Collaboration. We have not included ‘clinical’ RCTs but have included ‘health’ related RCTs as consistently as possible. This is because RCTs are now required for testing of pharmaceuticals, and clinical RCTs are much more advanced in Australia and around the world.

<sup>44</sup> “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 90.

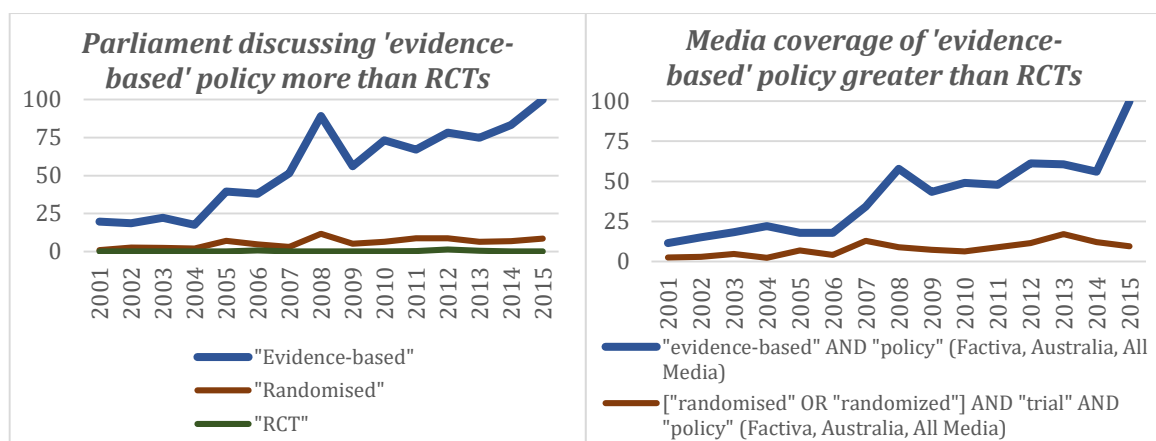


Figure 4 - Analysis of discussion of evidence-based policy and RCTs in Federal Parliament and Australian media<sup>45</sup>

One reason for this may be that discussions of ‘evidence-based’ options in Australia often appear to rapidly turn to the weaknesses and shortcomings of RCTs. For example, the Chief Economist in the federal Department of Industry, Innovation and Science recently published guidance on impact evaluation. This included numerous descriptions of where RCTs would not be appropriate, and concluded by saying that “calls for more use of RCTs ... [are] understandable—everyone wants simple answers to questions, but as described above RCTs are not always possible or useful.”<sup>46</sup>

Even the NSW Department of Family and Community Services describes in their *Behavioural Insights Toolkit* that “In the context of FACS, RCTs may not be possible... Therefore.... it is appropriate to consider how other types of evaluation methods (such as before and after evaluation) could be used instead.”<sup>47</sup> This is despite many of Australia’s globally-relevant past social policy RCTs having taken place within FACS’ own policy domains, such as parenting.<sup>48</sup>

We don’t dispute the limitations to RCTs outlined in these documents, or even the wisdom of the guidance. Instead, we suggest that within the constraints of RCTs, when properly construed, there remain unrealised opportunities for their increased use in Australian governments – as demonstrated abroad. Accordingly, the Australian conversational pivot from evidence-based policy to the weaknesses of RCTs does not do justice to the potential value RCTs can bring when applied appropriately.

45 Hansard analysis based on Federal Parliament Sessions with Mentions of terms based on searches of Hansard, all chambers and committees, standardised (to adjust for increase overall volume) and indexed to 100). Media analysis based on Factiva searches of Australia, All Media, standardised (to adjust for increase overall volume) and indexed to 100. Sources: “Search Hansard – Parliament of Australia”; “Factiva.com.”

46 Rogers et al., “Choosing Appropriate Designs and Methods for Impact Evaluation,” 53.

47 “NSW FACS Behavioural Insights Toolkit,” 4.

48 See *Appendix Australian RCTs in Social Policy* for a list of RCTs in Australia, many of which look at families and community services



## 2. Start RCTs Now

### ***Recommendations***

Senior public servants should start by commissioning small, uncontroversial, high impact/cost likelihood, service optimisation RCTs. The following are examples that have been successful from around the world that should be considered:

- Moving the signature box to the top of forms involving disclosure to government, for example using the phrase “I promise that the information I am providing is true and accurate”.
- Increasing collection of owed payments to government using behaviourally-informed letters, for example identifying the social norm of making such payments on time.
- Increasing payment of fines using behaviourally-informed SMS reminders, for example sending easy, convenient messages prior to a bailiff being sent to an individual’s home.

Ministers should start pursuing an RCT in social policy. Start with RCTs that are likely to be successful, and start in areas likely to provide value for money for government. To maximise chances of success, consider one of the following globally successful, high value/cost interventions from around the world:

- Cognitive behavioural therapy to reduce crime – for example the *Becoming a Man* program which used “impulse control, future orientation and an understanding of proper conflict resolution” to reduce violent crime by 44% among young males.<sup>49</sup>
- Meta-cognition and self-regulation programme to improve educational outcomes – for example the UK Education Endowment Fund assessed that such programs averaged eight months’ additional progress per pupil for only £80 per student.<sup>50</sup>

We describe below that the best way to pursue RCTs in government is to start trying. Learning from experience and working on actual, current RCTs we believe will lead to long-term RCT capability being developed. It will provide the necessary context for our recommendations relating to culture and leadership, expertise and transparency. Which interventions should be trialled will depend on the department and the context. Nonetheless, to mitigate the risk of null results for these first RCTs, we outline below some options for interventions that could be considered to maximise the odds of initial success by focusing on what has worked elsewhere numerous times.

Both optimisation RCTs and policy RCTs can be very valuable to governments. RCTs focused on optimisation are typically less impactful, but much cheaper and more viable. RCTs focused on policies and programmes typically have the capacity to be more valuable (in terms of government effectiveness), but are riskier, often costlier and face more hurdles.

David Halpern, CEO of BIT, has defended emphasising optimisation RCTs through the idea of ‘radical incrementalism’, which he describes as “The idea that dramatic improvements can be achieved, and are more likely to be achieved, by systematically testing small variations in everything we do, rather than through dramatic leaps in the dark.”<sup>51</sup> This approach was criticised by *The Economist* in 2015 through an analysis of the over 100 RCTs at the UK Education Endowment Foundation:

---

49 Heller et al., “Thinking, Fast and Slow?,” 2015.

50 “Meta-Cognition and Self-Regulation | Toolkit Strand | Education Endowment Foundation,” -.

51 Halpern, *Inside the Nudge Unit*, 291.



*“[They] deal with small-bore questions, such as whether teenagers learn more if the school day starts later. Meanwhile the government is radically reshaping the management and funding of schools nationwide—without testing the changes first, let alone running trials. That is reckless.”<sup>52</sup>*

We think governments should do both. There is value in the application of both optimisation and policy RCTs, and no reason to only focus on one or the other. We recommend Australian governments work to normalise both the use of RCTs to optimise existing policies and to incorporate RCTs into new policies and programs.

This approach allows public servants to transition through the stages of learning Barber describes as relevant to public sector change: from unconscious incompetence through conscious incompetence and conscious competence to unconscious competence.<sup>53</sup> Early, small RCTs help public servants begin to appreciate the potential of RCTs, how much they could be used in the future, and how they contrast with many existing approaches to policy evidence. This approach of starting small to build familiarity was cited as a key to the growth and success of the UK’s Behavioural Insights Team. The early results from BIT were critical to showing “people across government that trials could be rapid and low-cost.”<sup>54</sup>

In addition, our survey of Australian Parliamentarians revealed that of those parliamentarians who had used RCTs to inform their policy decisions in the past, 67% chose RCTs as one of their top three priority policy inputs for the future (see graph on right). In contrast, only 12% of parliamentarians who had not used RCTs before prioritised them for the future. This reveals the power of familiarity with RCTs. Simply starting to use RCTs should see more demand for them in the future. Those whose departments or agencies have already started should proceed to build their RCT capabilities more broadly.

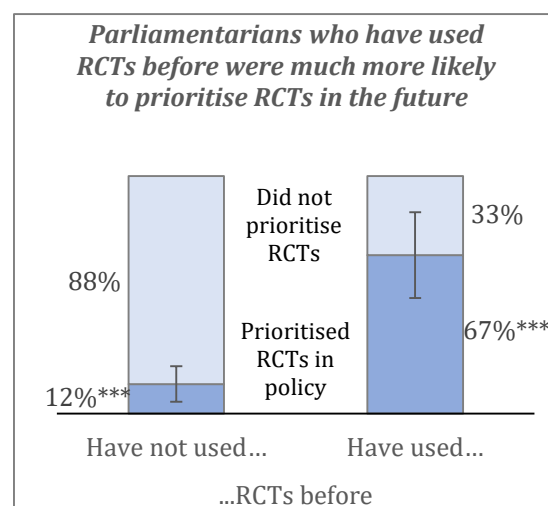


Figure 5 - Results from Survey of Australian Parliamentarians

## 2.1. Specific recommended interventions for early RCTs

### ***Service optimisation interventions for RCTs***

*Move the signature box to the top of forms involving disclosure to government*

Research has found that putting a signature box at the top of a form increases the honesty with which that form is completed.<sup>55</sup> This has been replicated in a range of contexts. Recently the US Social and Behavioral Sciences Team ran an RCT to increase the honesty of self-reported quarterly transactions by vendors paying an industrial funding fee.<sup>56</sup> The randomly assigned group, who signed at the top of the page “I promise that the information I am providing is true and accurate” before completing the form, reported on average \$445 more sales than the control group. In just the third quarter of 2014 this change increased remittances by \$1.59 million, merely as a result of moving the signature box to the top.

<sup>52</sup> “In Praise of Human Guinea Pigs.”

<sup>53</sup> Barber, *How to Run a Government*, 2015, 145.

<sup>54</sup> Halpern, *Inside the Nudge Unit*, 274.

<sup>55</sup> Shu et al., “Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End.”

<sup>56</sup> “Social and Behavioral Sciences Team Annual Report,” September 2015.

### *Increase collection of owed payments to government using behaviourally-informed letters*

Many governments around the world have increased revenue by using RCTs to improve how they collect payments that they are owed. The UK government designed a behaviourally-informed tax letter for those yet to pay their taxes, which informed recipients that others in their community had already paid. This increased compliance rates by up to 15 percentage points, freeing up an estimated £30m annually.<sup>57</sup>

### *Increase payment of fines using behaviourally-informed SMS reminders*

In the UK, BIT ran a successful RCT using behaviourally-informed SMS reminders to encourage people to pay court fines prior to a bailiff being sent to their homes. A number of interventions were trialled, all substantially improving payment rates. Sending these SMS reminders was estimated at being worth £860k per week to the national government.<sup>58</sup> The World Bank cites 73 papers across 6 domains where such text reminders have been used effectively.<sup>59</sup>

## ***Social policy interventions for RCTs***

Both of the interventions outlined below have worked in numerous settings elsewhere, rely on data that exists in administrative datasets for many governments, and can provide substantial benefits.

### *Crime – Cognitive Behavioural Therapy (CBT)*

CBT focusses on teaching offenders the cognitive processes and choices related to criminal behaviour, and how to restructure their thinking. An example of this is the *Becoming a Man* program from Chicago, which “seeks to help youth develop coping skills for managing situations that might otherwise lead to violence and other negative outcome”.<sup>60</sup> A review of 58 RCTs of CBT found an average 25% reduction in reoffenders in the intervention group in comparison to the control group.<sup>61</sup> The cost of this intervention ranges from \$80 per participant to \$2000 per participant depending on the resource intensity of the intervention.<sup>62</sup>

### *Education – Meta-cognition and self-regulation program*

Identified by EEF as one of the most effective and lowest cost interventions in education, meta-cognition and self-regulation approaches “help learners think about their own learning more explicitly”.<sup>63</sup> At an estimated cost of merely £80 per pupil, this approach has achieved an average “eight months’ additional progress” per pupil.<sup>64</sup>

---

57 “Applying Behavioural Insights to Reduce Fraud, Error and Debt,” 22.

58 Haynes et al., “Collection of Delinquent Fines.”

59 “Mind, Society, and Behavior - World Development Report,” 197.

60 “Winning Entry: Becoming a Man (B.A.M.).”

61 “Cognitive Behavioural Therapy (CBT).”

62 Heller et al., “Thinking, Fast and Slow?,” 2015, 36.

63 “Meta-Cognition and Self-Regulation.”

64 Ibid.

## 2.2.Criteria for which RCTs to pursue

RCTs have the potential to affordably and ethically provide valuable insight into the effectiveness of government activity. However, there's no guarantee RCTs will always be effective. To avoid these pitfalls, and strategically focus efforts to increase the utilisation of RCTs, we recommend policy makers seek:

- 1. Good data** – Can you accurately measure the outcome of interest using existing data?
- 2. Suitable populations** – Can you, in a fair way, randomise a reasonably large population?
- 3. Strategic interventions** – Are interventions standardised, easily varied, and impactful? <sup>65</sup>

These criteria are not exhaustive, but rather they are the three critical initial considerations for governments. They align with the description in the report discussed above by the Australian Department of Industry, Innovation and Science, that RCTs require that “evaluators can define the intervention in such a way that what was tested could be reproduced; ... random allocation into treatment and control groups; and that the sample size is sufficient to detect differences between treatment and control group.”<sup>66</sup>

We recommend these criteria because, as summarised by Cobb-Clark, “Better than nothing is not the same as good enough.”<sup>67</sup> There is a cost to running poor or inappropriate evaluations and trials, and so the criteria for when to use which type are important. We recommend starting with these three criteria for using RCTs. There are more, but these are the minimum hurdles.

### 2.2.1. Good data– Can you accurately measure the outcome using existing data?

Some policy areas are rich in existing data about the outcomes of most interest to policy makers (e.g. tax, health and education), while others have much weaker existing datasets (e.g. homelessness). Collecting new data, particularly using surveys, can be expensive and challenging. Instead, while building familiarity with RCTs, start where the data already exists as it will be cheaper and easier to confirm quality of data. This means there is great value to existing administrative datasets, particularly when they are linked with each other within and across jurisdictions.

For example, the first trial of the UK ‘nudge unit’ with the UK tax collection agency (HMRC) was considered “ideal” because the government “didn’t need to set up an expensive separate measurement system to test whether a letter worded one way or another was more or less effective, since HMRC already had systems in place to track who responded and paid.”<sup>68</sup> Similarly, Jason McNamara from Prime Minister & Cabinet’s Behavioural Economic Team of Australia (BETA) has described their approach as starting in “areas where there aren’t data issues”, before looking to scale.<sup>69</sup>

At a strategic level, Todd Rogers, Associate Professor of Public Policy at the Harvard Kennedy School who has run over 100 RCTs, recommends that the “ideal outcome measure can be acquired at no marginal cost, collected at the level of the target unit... comprehensively measured for all targeted units, and available rapidly after the experiment is completed.”<sup>70</sup> To achieve this, many of the recent wave of RCTs have been either utilising data that already exists in one place, or have

---

<sup>65</sup> These criteria build from the recommendation of Rogers, “Low Cost/ High ROI Experiments in Organizations.”

<sup>66</sup> Rogers et al., “Choosing Appropriate Designs and Methods for Impact Evaluation,” 8.

<sup>67</sup> “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 84.

<sup>68</sup> Halpern, *Inside the Nudge Unit*, 274.

<sup>69</sup> McNamara, interview, n.d.

<sup>70</sup> Rogers, “Low Cost/ High ROI Experiments in Organizations,” 5.

linked data that already exists. We propose the following scale for thinking about what outcomes measurement to attempt:

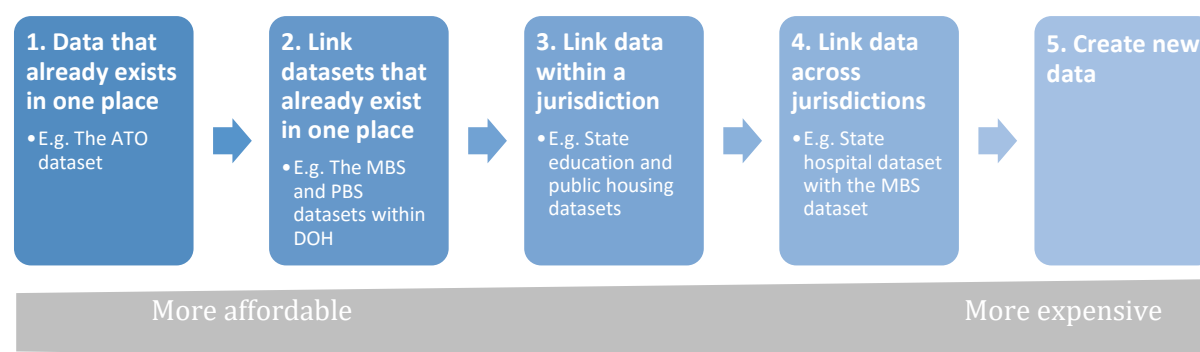


Figure 6 - Spectrum of data for use in RCTs

As Rogers recommends, “ideally, the outcome should already be measured and collected, which makes the marginal cost of using it in a randomised experiment trivial.”<sup>71</sup>

With this in mind, the Laura and John Arnold Foundation announced in 2015 funding for low-cost RCTs, specifically recommending that researchers use existing administrative data to measure key outcomes.<sup>72</sup> They assert that “researchers can significantly reduce the cost of evaluation by using these types of data rather than conducting expensive original data collection efforts such as interviews or tests.”<sup>73</sup>

## 2.2.2. Suitable populations – Can you, in a fair way, randomise a reasonably large population?

Small RCTs risk struggling for statistical power. Accordingly, as shown in Figure 8 to the right, most RCTs on the American Economic Association’s (AEA) Social Science RCT Registry have between 1,000 and 10,000 participants, and 67% of RCTs have over 1,000 participants.<sup>74</sup> While RCTs can work with smaller populations, starting with larger populations will reduce the risk of sample size concerns.

An intuitive option for increasing sample size is to design interventions that work with individuals (e.g. students, employees) instead of groups of individuals (e.g. schools, companies), since it is much more difficult to implement interventions across many large groups. Analysis of the AEA Registry reveals that 46% of RCTs were randomised at the level of the individual, compared to 8% by household and 19% by geography. However, it is not necessary to only work with individuals in order to avoid sample size issues. For example, many of the RCTs from the EEF in the UK have been at the school level, often with reasonably small sample sizes. Analysis of the EEF’s RCTs reveals that more than 90% of their trials involved randomisation of 150 schools or fewer. For more detail on what this means in practice, see the Appendix, Table 10, which outlines the minimum required sample sizes for a range of possible RCTs.

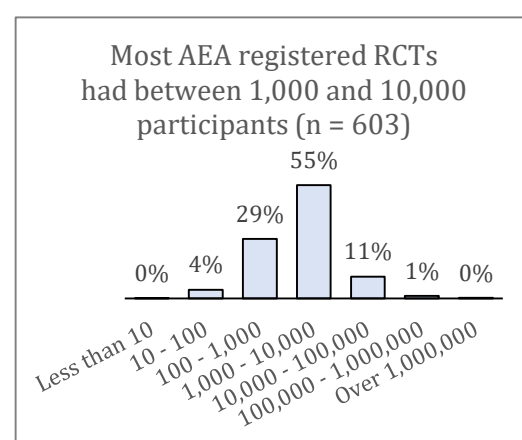


Figure 7 - Sample sizes of RCTs on AEA RCT Registry

<sup>71</sup> Ibid.

<sup>72</sup> “Request for Proposals: Low-Cost Randomised Controlled Trials to Drive Effective Social Spending,” 8.

<sup>73</sup> “Laura and John Arnold Foundation Announces Expanded Funding for Low-Cost Randomised Controlled Trials to Drive Effective Social Spending - Laura and John Arnold Foundation.”

<sup>74</sup> “AEA RCT Registry.”

In addition, policymakers must make sure they have a population that can be randomised to receive or to not receive the intervention in a fair way. Some interventions, even if they are very effective, are so minor in nature that it is unlikely to be considered unfair if some people do not receive them (e.g. stickers on letters, or an additional graphic in an email). Other contexts, such as life-saving emergency responses, may be considered out-of-bounds for random allocation to interventions. However, substantial interventions often do present opportunities for fair random allocation. For example, if there is more demand than supply for a new service, lottery-based access may be the fairest process. Alternatively, a randomised staggered rollout can ensure that all participants access the intervention in due course, but with different start dates. While it is certainly important to consider the fairness implications of having a control group, we suggest it is unhelpful to argue that control groups are unethical *prima facie*. As Halpern asks, using a comparison to the spread of RCTs in medicine, “is it really credible to say that systematic testing of medical treatments is ethically acceptable where the outcomes are measured in life and death, but that such methods are not to be used to test the efficacy of welfare or education?”<sup>75</sup>

### 2.2.3. **Strategic interventions – Are interventions standardised, easily varied, and impactful?**

To test the effectiveness of an intervention, RCTs need interventions that are the same throughout the trial. Interventions that are difficult to standardise are inherently difficult to measure for effectiveness (both for RCTs and beyond). Some interventions will be easily varied (e.g. the content of some newly created webpages), and some are very difficult to vary (e.g. policy levers subject to detailed, multi-party regulations). Start with the interventions that are standardised and easily varied. Further, use existing evidence from around the world to focus on the interventions that have the highest likelihood of making the biggest difference. *Figure 8* outlines a case study of such an RCT in the UK.<sup>76</sup>

#### **Standardised**

The most attractive element of an RCT is its ability to attribute causation to a given, specific intervention. For that to be possible, the interventions being tested need to be constant across all participants in the treatment group, and held consistent for the duration of the RCT. If the intervention varies within a treatment group, it will be impossible to identify both what the intervention actually is, and whether it worked. Further, if interventions are adapted and varied as the trial goes, it will be impossible to evaluate which parts of the intervention worked, and which didn't. As outlined by Luca & Hauser, “Once your experiment is running, leave it alone!”<sup>77</sup> The need to have standardised interventions once more highlights the attractiveness of RCTs in the digital context. Online material such as emails, webpages and processes can be easily designed and locked in for a given trial period.

#### **Case Study**

In 2013 the UK government ran an online RCT to increase organ donation registrations in a partnership across the Behavioural Insights Team, Cabinet Office, the National Health Service (NHS), Department of Health, Driver & Vehicle Licensing Authority and the Government Digital Service. The trial tested the impact of eight behaviourally-informed webpages which appeared after citizens completed renewing their vehicle tax or registering for a driver's license.

In the five-week trial, they tested very minor changes to the information displayed on a web page. Over 1 million people visited the page, 135,000 per variant. The results of the RCT showed that 6 out of the 7 behaviourally-informed pages outperformed the control page, and one underperformed the control. The ‘Reciprocity’ page, which simply included the sentence “*If you needed an organ transplant would you have one? If so, please help others*” led to 39% more registrations than the control (statistically significant at 0.001). This is the equivalent of 100,000 additional registrants per year as a result of adding one sentence to a webpage!

*Figure 8 - Organ donor registrations in the UK*

<sup>75</sup> Halpern, *Inside the Nudge Unit*, 273.

<sup>76</sup> Harper, “Applying Behavioural Insights to Organ Donation: Preliminary Results from a Randomised Controlled Trial.”

<sup>77</sup> Hauser and Luca, “How to Design (and Analyze) a Business Experiment.”

That is not to say that tougher, more complex interventions in social policy cannot work with RCTs, but rather that interventions need to be standardised and held steady for RCTs to be informative. An example of such an approach outside of the digital context is the Multidimensional Treatment Foster Care (MTFC) licensing and monitoring programme to ensure fidelity to the intervention. MTFC is an intervention that supports teenagers with chronic or severe behaviour problems. This intervention has proven to be highly effective in many different contexts. In order to test and maintain the effectiveness of the program, the developers of MTFC (Treatment Foster Care Oregon)<sup>78</sup> implemented a thorough monitoring approach to ensure fidelity to the programme design by all providers around the world. Their fidelity monitoring includes weekly consultations, video recordings, site visits, programme assessments every 9-12 months, and expert consultants.

### ***Cheaply and easily varied***

The second consideration is to look for interventions that can be varied cheaply. Rogers recommends thinking about the “vector” through which the intervention is delivered.<sup>79</sup> For example, easily amendable vectors might include webpages, emails and SMS. In some organisations the vector of letters may be easily amended, while in some bureaucracies it might take enormous cost and effort to change a letter. As a rule, “the more people who are needed to develop, manage, and monitor the administration of different variations of the vector, the costlier the experiment will be and the less likely it is that it will be implemented as intended.”<sup>80</sup>

We identified two drivers of when a “vector” is easily varied: *infrastructure* and *people*. While in some places with capable, enthusiastic people, ageing infrastructure may mean it takes up to 18 months to change the content of a letter, other places may have cutting edge technology but a lack of interest or willingness to participate. This means it will be important to start in those places where both the infrastructure and people allow alternative interventions to be trialled cheaply.

We would generally expect policy RCTs to be more expensive than optimisation RCTs. However, that does not mean policy RCTs always involve highly expensive interventions. Analysis of the EEF’s first 119 RCTs in education reveals that 28% of trials cost less than £250k (for both intervention and evaluation costs) and 55% cost between £250k and £1m. While these costs are higher than the USD\$100k - \$300k “low-cost RCT” approach advocated by the Arnold Foundation,<sup>81</sup> in the context of education budgets they are still reasonably low.

### ***Impactful***

The third consideration for which interventions are suited to RCTs is to look for interventions expected to make the biggest difference. The bigger the “effect size” the more likely RCTs will produce statistically significant results. Hauser & Luca (2015) advocate using exaggerated mechanisms to increase effect size, calling this approach “Use a big hammer”!<sup>82</sup> This consideration should be weighed against the cost and ease of variation. While adding a new sentence to a webpage may be a consistent, cheaply varied intervention - like BIT did for organ donation - there may have been alternative interventions with a greater effect size but which were less consistent and affordable. It is possible that having a call centre ask residents to join the organ donor registry would have had a larger effect, and so while more expensive and less standardised, this approach would be more attractive on this dimension.

---

78 “Treatment Foster Care Oregon.”

79 Rogers, “Low Cost/ High ROI Experiments in Organizations,” 4.

80 Ibid.

81 “Laura and John Arnold Foundation Announces Expanded Funding for Low-Cost Randomised Controlled Trials to Drive Effective Social Spending - Laura and John Arnold Foundation.”

82 Hauser and Luca, “How to Design (and Analyze) a Business Experiment.”



### 3. Build RCT Capability

Once the government has begun building experience with RCTs, there is the larger task of establishing a testing culture across government. In the first instance, this will require repetition – don't stop after one RCT. As can be seen from *Appendix 4.3 Australian RCTs in Social Policy* there have been policy RCTs in the past in Australia, many with government partners, but it appears that those initial efforts were not translated into momentum for RCTs in policy more broadly.

Instead, policymakers should learn from their initial RCTs, and then develop a portfolio of RCTs using a range of interventions. A portfolio is necessary because not all RCTs will demonstrate positive effects. Rather than being challenged by the notion that “some things are shown to not work when they're rigorously evaluated,”<sup>83</sup> policymakers need to embrace both positive and negative findings from RCTs, since a portfolio of results will be required for governments to effectively shift their priorities and choices away from what is not working to what is working.

As the government's capability for RCTs develops, there are two attractive types of large-scale policy RCTs to consider:

#### ***1. Where there is more demand than supply, use lottery-based access***

Many now argue that where access to a programme is limited, lotteries may be the fairest way to determine who gets access. Lotteries reduce the potential for explicit or systemic bias to influence any selection criteria, and have the benefit of being a transparent, consistent decision rule. Of course, in addition, it allows governments to learn much more about the actual effectiveness of the program. There are numerous examples of governments and the social sector using lotteries in these contexts, including Moving to Opportunity, the Vietnam Draft Lottery,<sup>84</sup> KIPP Charter Schools,<sup>85</sup> and PROGRESA (conditional cash transfer program) in Mexico.<sup>86</sup>

One famous example of this approach, the Oregon Health Insurance Experiment, is outlined in the case study in *Figure 9*.<sup>87</sup> This sort of insight into a nationally relevant, high cost programme is rare, valuable, and only possible because the programme was expanded through a lottery.

#### ***2. In universal coverage, use step-wise rollouts***

For programmes that are to be rolled-out to entire populations, an option is to randomly rollout the programme. This allows for all individuals to access the programme in time, and for the government to learn whether the programme is

#### **Case Study**

In 2008, Oregon decided to expand access to Medicaid, but could not afford to admit the entire waitlist. They decided that the fairest selection mechanism was to use a lottery. The state government partnered with academics to ensure the lottery design allowed for learning about the value of the programme to the health of citizens. From 75,000 people on the waitlist, around 10,000 randomly chosen people joined the programme.

Combining administrative data with surveys, health tests and interviews, this experiment provided invaluable insights into the actual causal (as opposed to correlational) impact of health insurance on people's lives. The results of the experiment include that Medicaid coverage significantly increased the probability of a diabetes diagnosis, increased the use of diabetes medication, decreased the probability of a positive screening for depression, increased the use of a range of preventative services and decreased (nearly eliminated) catastrophic out-of-pocket medical expenses.

*Figure 9 - Oregon Health Insurance Experiment*

83 Baron, interview.

84 Hearst, Newman, and Hulley, “Delayed Effects of the Military Draft on Mortality.”

85 “Frequently Asked Questions.”

86 Gertler, “The Impact of Progres on Health; Final Report.”

87 Finkelstein and Taubman, “Using Randomised Evaluations to Improve the Efficiency of US Healthcare Delivery,” 30; Baicker et al., “The Oregon Experiment — Effects of Medicaid on Clinical Outcomes.”

effective and good value for money. An example of this approach is the Back-to-Work programme run by BIT and the UK Department of Work and Pensions.<sup>88</sup>

BIT had trialled an intervention in a Jobcentre in Loughton, Essex to improve the process of supporting people back into the workforce through increasing the use of ‘implementation intentions’ in job-search activities. The initial trial found a 5 percentage-point increase in people leaving benefits (“off-flow rates”) than the control group. To continue testing the efficacy of this intervention as the approach was scaled to the next 12 sites, they utilised a “step wedge” trial where all Jobcentres in Essex delivered the intervention in a randomly prescribed sequence so that by the end of 10 months all sites were delivering the intervention.

This allowed for both universal delivery of this intervention in this area, and randomisation to allow for an RCT-based evaluation. The results of this rollout with 110,838 participants saw off-flow rates in the treatment condition of 58.5%, up from 56.8% in the control condition (significant at 0.01). For a cheap, simple intervention, this is substantial value-for-money. While a smaller effect than the initial trial, it shows that it is possible to scale up interventions while randomising to test for effectiveness.

---

Opportunities for RCTs are everywhere. We see the three critical areas to seizing this opportunity as: culture and leadership, expertise and transparency. The following section outlines what we view to be necessary changes in each of these areas.

## 3.1. Culture & Leadership

### **Recommendations**

Ministers should demand evidence and RCT evaluation strategies from public servants and allocate funding based on evidence.

Senior public servants should push available evidence to ministers and outline the most rigorous possible evaluation strategies for new policies.

Our interviews with practitioners and experts were near-unanimous in claiming that the biggest barrier to more evidence-based policy in Australia is political culture and leadership.<sup>89</sup> RCTs require a culture willing to acknowledge that we do not know whether policies or programmes actually work. Halpern describes “our dangerous tendency to overconfidence and our presumption that what we do know is ‘right’.”<sup>90</sup> Instead, leaders needed to “get used to saying ‘I don’t know – but I know how we can find out.’ We can test, learn and adapt. And we can do it fast.”<sup>91</sup>

Developing a culture with the humility to support RCTs will require action from both ministers and public servants. Our view is that Australian governments are presently taking too few smart risks. Shergold summarises the problem as:

*Taking the first step requires trust. Departments must trust that their ministers will back them, so that they can learn from mistakes. Ministers must trust that citizens have the common sense to see that it is smarter to ‘fail fast, fail small’ rather than to pretend that*

---

<sup>88</sup> “The Behavioural Insights Team, Update Report 2013-2015,” 7.

<sup>89</sup> We interviewed more than 20 practitioners in and outside of government across the UK, US and Australia.

<sup>90</sup> Halpern, *Inside the Nudge Unit*, 297.

<sup>91</sup> *Ibid.*, 268.



*failure is impossible. Citizens must trust that government will learn and improve, and that an unsuccessful trial is not a waste of public resources. Being agile needs to be authorised.*<sup>92</sup>

There is an opportunity to provide this political leadership in Australia. Cobb-Clark argues that “we are waiting for someone with real leadership ability”.<sup>93</sup> She cites a “general culture of not making policy on the basis of strong evidence” as a key barrier in preventing rigorous evaluation from spreading in Australia.<sup>94</sup> This view was supported by Burtless, who observed that “many Australian officials and researchers deeply oppose the most reliable kind of programme evaluation – randomised social experiments.”<sup>95</sup> Shergold reflected that trials have too often been used “to fob off interest groups or to avoid committing the level of resources necessary to tackle a hard problem.”<sup>96</sup>

Abroad, political leadership was also cited as critical. Support from Presidents Bush and Obama was instrumental in issuing directives to make the federal government more evidence-based. In the UK, the access and permission of the Prime Minister’s office was crucial to establishing BIT. Similarly, the Social Impact Bond Lab based at Harvard Kennedy School (SIB Lab), which advises governments around the USA on Pay for Success schemes, includes in its strategy as the first step, before anything else, to ensure “enthusiasm and commitment among leadership”.<sup>97</sup>

Our survey results confirm that there is an opportunity in Australia to show leadership. 73% of respondents support the use of RCTs in social policy. And of those politicians who have used RCTs to inform their policy decisions in the past, 67% chose RCTs as one of their top three priority policy inputs for the future. This demonstrates that a bold politician seizing the opportunity to use RCTs may well encounter support and have a high value assigned to the evidence produced.

### 3.1.1. Institutional measures

#### **Recommendations**

**Both Ministers and senior public servants** should request that each year every agency identify:

- Three policy areas with limited evidence of effectiveness
- Actions to be taken to increase the evidence base
- At least one RCT they intend to run
- Results from the previous year’s RCT(s)

Requiring agencies and departments to improve evidence of policy effectiveness focuses resources on the challenge. The Obama Administration asked agencies to submit 2-3 programme areas to OMB they were going to improve their evaluation on over the next 12 months.<sup>98</sup> Subsequently, OMB allocated \$100m to support 35 rigorous programme evaluations and capacity-building projects. This forcing mechanism helps departments and agencies to focus on improving the quality of their evaluation.

Ron Haskins summarises the approach OMB took in his book *Show Me the Evidence* in three steps:

1. New funding for social programmes “would go primarily, but not exclusively, to programmes that showed strong evidence of success”;<sup>99</sup>

---

<sup>92</sup> Shergold, “Learning from Failure,” 80.

<sup>93</sup> Cobb-Clark, interview, November 8, 2015.

<sup>94</sup> Ibid.

<sup>95</sup> Burtless, “Innovations in Labour Market Policies,” 100.

<sup>96</sup> Shergold, “Learning from Failure,” 80–1.

<sup>97</sup> Liebman and Sellman, “Social Impact Bonds, A Guide for State and Local Governments.”

<sup>98</sup> Orszag, “Evaluating Programs for Efficacy and Cost Efficiency,” 1.

<sup>99</sup> Haskins and Margolis, *Show Me the Evidence*, 2015, 6.

2. Programmes would be implemented at the state and local level and subject to continuous evaluation; and;
3. Over time, all federally-funded programmes would be required to show evidence of their efficacy.<sup>100</sup>

This approach “makes room for less than RCT evidence” and is focused on increasing the evaluation of all federal spending, regardless of method.<sup>101</sup> Taking OMB’s approach of reporting on evaluation will increase the priority of improving evaluation for Australia’s agencies and departments.

To take this opportunity, evaluators must be included in early policy discussions. Cobb-Clark notes that too often in Australia “the evaluator is not called in until the project is already well advanced, and there is a tight deadline for completing the evaluation, frequently combined with a limited budget and without access to baseline data.”<sup>102</sup> The US federal government sought to overcome this problem by holding workshops on rigorous evaluation where someone from each department’s policy and evaluation teams attended.<sup>103</sup> These workshops covered:

- How can agencies focus evaluation resources on the most important programme and policy questions?
- How can agencies use administrative data sets from multiple programs and levels of government to answer important questions while protecting privacy?
- How can agencies conduct rigorous programme evaluations and data analytics on a tight budget?
- How can agencies use their existing authority to turn a traditional competitive grant programme into an innovative, evidence-based one?
- How can agencies harness research findings from the social and behavioural sciences to implement low-cost approaches to improving programme results?

Evaluators need to be present from the start, otherwise opportunities to randomise will be missed. A similar workshop series could be run by the Australian government to build its capacity to enforce the recommendation to improve evaluation in three policy areas each year.

### 3.1.2. Communication strategies

In discussing how he talks about trials and behavioural science in the UK government, David Halpern summarises: “If you really want to achieve impact on a large scale... it’s *conversion* not *compliance* that you’re after. For conversion, you need to persuade and convince, not force and insist.”<sup>104</sup>

Many people we spoke with discussed the importance of not beginning conversations about evidence by talking about RCTs themselves. RCT education starts with education about evaluation in general and the importance of finding out what works. This involves starting with the reality that lots of the things people think work, turn out not to. As Jon Baron describes, “Don’t lead with the methodology, lead with the need.”<sup>105</sup>

RCTs do have detractors, and communicating their benefits is not always straightforward. Public perception may not initially favour RCTs, but this attitude can shift; as Shergold explains, “announcing a trial need not communicate a lack of commitment to following through on a policy.

---

<sup>100</sup> Ibid.

<sup>101</sup> Ibid., 20.

<sup>102</sup> “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 84–5.

<sup>103</sup> Burwell et al, “Next Steps on the Evidence and Innovation Agenda”. 4.

<sup>104</sup> Halpern, *Inside the Nudge Unit*, 274.

<sup>105</sup> Baron, interview.

On the contrary, it should mean that government is so committed to achieving a successful outcome that it will carefully investigate the best way of doing so.”<sup>106</sup>

Even then, many objections to the use of RCTs in social policy may be raised. Stuart Buck and Josh McGee from the Arnold Foundation have documented the common objections in their article *Why Government Needs More RCTs – Refuting the Myths*, which are summarised and built on in the table below. These responses are polemical, and assume that the RCT meets the criteria outlined above.

### ***Common RCT Talking Points***

<b>Concern</b>	<b>Explanation</b>	<b>Response</b>
‘RCTs are expensive and slow’	RCTs often track long-term outcomes and require one-off data collection. This makes RCTs infeasible to use frequently to evaluate government programmes.	RCTs have been expensive in the past, but have become cheaper and faster. Collecting new data on any programme participants can be expensive, but that is true for any type of evaluation. Where possible, data that the government already collects should be used, as randomisation itself is not expensive.
‘RCTs are unethical’	Treating citizens differentially through randomisation is unethical. Moreover, it is unethical to deny the control group an intervention with a high-likelihood of success.	Elizabeth Linos, from the US arm of BIT, contends that “it is important to clarify that we take ethics seriously, and mean it – not just to persuade.” <sup>107</sup>  It is exceedingly rare that an intervention will so obviously work that evaluation is not required. This approach can lead to disastrous results, such as the ‘Scared Straight’ programme to deter juvenile delinquency, which non-random evaluations showed worked but which was actually shown to increase crime through a series of RCTs. <sup>108</sup>  As billions of dollars are spent on social programmes without knowing if they have any impact on outcomes, there is an ethical burden to pursue effectiveness. Often new social programmes are trialled before they are rolled out to the entire eligible population, and conducting this trial / rollout randomly can be fairer than other ways.
‘RCTs are narrow’	External validity is low in RCTs, reducing the generalisability of findings.	This can be true of both RCTs and other types of evaluation. However, it does not need to be true, due to the possibility of replicating RCTs in diverse settings to ensure the outcome improvement is robust in the different contexts.
‘RCTs are a black-box’	RCTs only demonstrate a change in outcomes, not which part of the intervention caused the change in outcome.	While this can be true, it again applies to many forms of evaluation. Moreover, RCTs can – through a factorial design – create different treatment arms that isolate which part of the treatment is affecting the outcome more effectively.
‘RCTs are not suited to complex, fast-	RCTs are too involved to be practical in many settings.	If a programme is changing too quickly for an RCT, it is likely to be changing too quickly to conduct any form of statistically significant quantitative evaluation.

<sup>106</sup> Shergold, “Learning from Failure,” 80–1.

<sup>107</sup> Linos, interview.

<sup>108</sup> “Test Learn Adapt,” col. Box 7.

Concern	Explanation	Response
changing programs'		However, most government activity and programmes do not fit this description.
'RCTs can still be biased'	RCTs are as susceptible as other evaluations to manipulated results.	No evaluation method is perfect, but RCTs by design are less vulnerable to bias than almost all other forms of evaluation. Particularly, observational studies have greater room for manipulation. This risk can be further mitigated by pre-registering the RCT on a public registry.

Table 4 - Responses to concerns about RCTs<sup>109</sup>

## 3.2. Expertise

### 3.2.1. Centre of Excellence

#### **Recommendations**

Ministers and senior public servants establish BETA as a 'centre of excellence' for RCTs which collaborates with departments to provide RCT expertise for the Federal Government

We recommend a central function in government to provide RCT expertise and promote quality evaluation. Barber argues that the role of a central team is threefold: to set the strategy, monitor performance and provide human capital.<sup>110</sup> Increasing the use of RCTs requires these three components, making a compelling case to create a unit at the centre of government. Moreover, Nicholas Gruen has advocated for an "Advocate for Government Innovation" whose remit would be to provide "funding and resources for randomised policy trials" in addition to:

- Facilitating innovation in government;
- Improving service delivery; and;
- Facilitating the regulatory needs of innovators.<sup>111</sup>

The functions that our proposed centre would fulfil are:

Centre Responsibility	Rationale
Set evaluation standards across government	A common set of processes for initiating and executing evaluations across government reduces the burden for individual departments. Moreover, it mitigates the likelihood of low-quality evaluations, which risk undermining the faith in all evaluations.
Promote and monitor evidence-based policy making	The recommendation that ministers and public servants require agencies and departments to identify policy areas with limited evidence of evaluation and then increase the evidence needs a central body to coordinate this work and monitor its implementation.
Provide a centralised hub of technical expertise and	Concentrating specialist evaluation skills that can then be "loaned" out when required will reduce the burden on individual

<sup>109</sup> Buck and McGee, "Why Government Needs More Randomized Controlled Trials: Refuting the Myths," July 2015.

<sup>110</sup> Barber, *How to Run a Government*, 2015, 92–93.

<sup>111</sup> Gruen, "Beyond Central Planning," 13.

coordinate resources departments and agencies	evaluation across and	departments and agencies to keep skills in-house. An example of this is the US Digital Services Team, which seconds out teams to departments for time-limited, specific projects. A central body can also “match-make” between academics and projects.
---	-----------------------	--

Table 5 – Proposed functions of a Centre of Excellence

There are three options that could fulfil these responsibilities:

1. *A central ‘centre for excellence’ housed in the Department of Prime Minister and Cabinet*

The centre would hire and second RCT experts to departments and agencies, as well as maintaining a network of experts who would be hired directly by agencies and departments. This approach has the benefit of standardising approaches to RCTs. It retains knowledge and builds capability within the public service. And because secondees are public servants there is likely to be more trust in the integrity of the advice from Ministers. This makes it more likely that secondees will be brought in earlier in the process. It may be the case that this is an additional function to the existing infrastructure of BETA, or another central team that already exists. McNamara of BETA described how BETA is already operating a version of this model, with a team of people who “go out and work with the partner agencies on their priorities.”<sup>112</sup>

2. *An independent agency to commission all policy evaluations on behalf of government*

Cobb-Clark advocates for an independent agency, set up in the manner of the Productivity Commission or Reserve Bank. This approach ensures the highest level of integrity of evaluations.<sup>113</sup> It would also serve as a registry for programme evaluations and provide technical expertise to government. As it resides outside of government, this option is less-able to provide technical expertise to government departments.

3. *The status quo*

There are pockets of expertise dispersed throughout the Australian Public Service (APS), and policymakers have the option to partner with academics or hire consultants as required. This option is the easiest to implement, but lacks a centre to promote RCTs across government.

<sup>112</sup> McNamara, interview, December 9, 2015.

<sup>113</sup> “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 90.

We assess these options against the following four criteria:

Criteria	Stronger option	Discussion
<i>Cost</i>	Status quo	The cost of each model depends on the intended size, and who pays depends on the option. However, the cheapest option is to persist with the status quo as additional resources would be required only as promising RCTs are identified.
<i>Technical quality</i>	Independent agency	A specialist, independent agency held at arms-length from government could arguably attract better talent. However, a 'centre for excellence' could achieve a similar outcome depending on its exact remit.
<i>Trust</i>	Centre	Ministers and the public service may be reticent in letting a fully independent agency into the most intimate policy settings. In contrast, a central team is likely to have a high degree of trust – as well as access.  Advice from the public service is free from commercial considerations, or the perception of them. This makes it more likely that their advice will be heard than relying on external consultants.  On the other hand, an independent agency may be better at building trust with stakeholders, such as the public and academics.
<i>Knowledge retention</i>	Centre	The long-term ability of the public service to conduct RCTs will be enhanced through experience. Part of the push for RCTs is to embed a 'learning culture', which is more likely when the centre is part of government. Using an independent agency or consultants means that learning resides outside of the APS.

*Table 6 – Assessment of options for centre of excellence*

For the Federal Government, BETA is best-placed to fulfil these responsibilities. 14 departments have agreed to fund and work with BETA, providing BETA with legitimacy and access.<sup>114</sup> Moreover, BETA has plans already in train to undertake the following:

- *Set evaluation standards across government* – BETA plans to establish an Internal Review Board (IRB) process for RCTs and set a norm of publicly and sensitively pre-registering trials
- *Promote and monitor evidence-based policy making* – Each partner department and agency to create workplans for 3-4 behaviourally-informed projects. These will be narrowed down in conjunction with BETA, with the intention of being run using an RCT for evaluation
- *Provide a centralised hub of technical expertise* – BETA seconds staff from partner departments to develop their behavioural and evaluation skills.

As BETA already exists with a very similar mandate, it is well-placed to fulfil the responsibilities of the 'centre for excellence'.

---

<sup>114</sup> Hiscox, interview.

### 3.2.2. Partnerships

#### **Recommendations**

Senior public servants should better engage the academic community and social sector.

#### ***Academic Partnerships***

There is a further opportunity to establish trusted relationships between the government and the academic community, in order to increase access to a great source of policy evaluation knowledge. In the UK, academics were central to the development of RCTs in Government. BIT has described how “with the right academic and policy support RCTs can be much cheaper and simpler to put in place than is often supposed.”<sup>115</sup>

Unfortunately, past experiences have taught some academics in Australia that the federal government is not serious about rigorous evaluation and so they do not want to invest their time again. Cobb-Clark argues that “the lack of a willingness to commit to eventual publication of results has meant that Australian academics are increasingly disengaged from evaluations of major economic and social initiatives.”<sup>116</sup>

Currently, academic publication plays a key role in quality assurance for policy RCTs around the world. However, some Australian public servants see academics as too self-interested in publishing their papers and less interested in creating programmes with great impact. There is a range of academic engagement options that could build more productive relationships. These options include:

<b>Academics to:</b>	<b>Pros</b>	<b>Cons</b>
<i>Use open data to evaluate programmes externally; separate from government</i>	Easy to do as relationships between academics and the public service do not need to be established	Unless there is a natural randomisation mechanism, this limits evaluation to observational studies. Moreover, as the evaluation is not commissioned by government it may not answer the questions necessary to improve the programme
<i>Provide consulting services on specific projects</i>	Academics engaged as partners on areas where they have particular knowledge	Academics may not be interested in being engaged purely to provide services unless the project may result in publishable research
<i>Second students to work on specific projects</i>	Large supply of students eager for work experience	Student knowledge and value may be limited; large coordination costs
<i>Partner with or embed themselves in a department to run a trial</i>	Academic attention dedicated close to full-time to a government-commissioned RCT	There may be a gap between the design and analysis necessary to create publishable research versus working out what works

*Table 7 – Assessment of options for partnering with academics*

We recommend the Australian government collaborate with academics through whichever models are feasible in each department’s policy context. Too much expertise resides in

<sup>115</sup> Haynes et al., “Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials,” 5.

<sup>116</sup> “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 90.



universities for academics to be overlooked. BETA is currently working to rebuild some of these links with the academic community. As Cobb-Clark surmises: “without serious, long-standing research partnerships between policymakers, academics, and other stakeholders, very little progress can be made.”<sup>117</sup>

### ***Social Sector Partnerships***

Similarly, there is an opportunity for government to partner with the social sector to design and run RCTs. In the US, many RCTs are funded by philanthropic foundations specifically pushing RCTs and evidence-based government. The interventions are frequently delivered by NGOs willing to embrace RCTs.<sup>118</sup>

For example, the Coalition for Evidence-Based Policy – now within the Arnold Foundation – played a key role in supporting OMB’s transition to requiring more evidence in policy.<sup>119</sup> Moreover, they fund low-cost RCTs and promote their use. To the extent that similar organisations exist or are set up within Australia, the government should look to partner with them.

Shergold makes the argument for closer partnerships between the government and the social sector, which frequently delivers services on behalf of the government:

*By collaborating with potential providers on policy design and execution, and paying on the basis of performance-based results, a greater spirit of partnership can be created... Perhaps the best role of the [government] is to stand aside and let the organisations commissioned to deliver the services get on with the job.*<sup>120</sup>

To achieve this vision, the government must demonstrate that it is willing to support the social sector. Through our conversations with the Gloria Gong, Executive Assistant Director at the Government Performance Lab at the Harvard Kennedy School we heard that the following is important for supporting the social sector:

- *Frame evaluation as a benefit – not something to hide from:* Mission-driven organisations benefit from evaluation because it allows them to better serve their population. Evidence of a successful intervention provides a compelling rationale for scaling the service, while evidence of no impact provides an internal case to pivot and try something new.
- *Support organisations willing to be subjected to rigorous evaluation:* RCTs frequently show poor results from an intervention. If a social sector organization volunteers for scrutiny, show them “huge respect. They are putting themselves under the microscope.”<sup>121</sup> This shows a dedication to learning and should not result in the severing of the relationship if the intervention is found to have little or no impact.<sup>122</sup>

---

117 Cobb-Clark, “Evidence-Based Policy: The Need for Data and Transparent Evaluations,” 30.

118 For example, “Laura and John Arnold Foundation Announces Expanded Funding for Low-Cost Randomised Controlled Trials to Drive Effective Social Spending - Laura and John Arnold Foundation.”

119 Baron, “Board of Advisors of the Coalition for Evidence-Based Policy.”

120 Shergold, “Learning from Failure,” 71.

121 Gong, interview.

122 Ibid.



## 3.3. Transparency

### 3.3.1. Public Registry

#### **Recommendations**

Ministers should authorise a public registry of policy evaluations.

A lack of evaluation transparency reduces the credibility of all evaluations. There is predictable cynicism about governments evaluating their own policies. One solution to the transparency deficit is to create a public registry of RCTs and policy evaluations. Cobb-Clark argues that greater transparency would:

- *Put pressure on evaluators to lift their game;*
- *Allow evaluations themselves to be evaluated against sound scientific principles so that we can make judgements about which to weight more heavily and which to ignore;*
- *Provide an opportunity for truly informed public debate about the issues facing us; and;*
- *Substantially enhance our chances for sound decision-making.*<sup>123</sup>

Publishing evaluation methodologies before the results are out reduces the scope for ‘massaging’ results. This will increase trust in the evidence produced by RCTs and evaluations.

This approach requires politicians to be prepared for policies to be shown to have little or no effect. As described above, analysis of the EEF’s database of RCTs in the UK revealed that of the 45 completed trials, only 11 had a positive, statistically significant result. The US SBST also reports null findings. Publishing these studies is as important as publishing studies with positive outcomes. Knowing what *not* to do can be as useful as knowing what to do. There is even now pressure on the UK government to be more public with the rest of its RCTs (beyond the EEF), with public interest NGO Sense About Science arguing that “we must now press for timely publication of all government research in the interests of accountable public policy.”<sup>124</sup>

Registering research before it is conducted also mitigates against the risk of researchers altering their methodology if their findings do not align to their expectations, a risk discussed by Buck and McGee.<sup>125</sup> Cobb-Clark has cited the example of the Australian and New Zealand Clinical Trials Registry for medical research as an example to be emulated for social policy evaluations.<sup>126</sup> She argues that “increased transparency and wider dissemination of results are absolutely essential to improving the quality and information content” of evaluations and policy.<sup>127</sup> Given the politically sensitive nature of many evaluations, a short embargo period may help assuage the fears of anxious politicians.<sup>128</sup>

The registry could be housed in a central agency, by a third-party or by expanding the mandate of an existing clinical trial registry. If it were housed in a central agency, it could be managed within the Department of Prime Minister and Cabinet, possibly by the BETA team. If it were outside of government, it could be hosted by a statutory authority, such as the Productivity Commission or a university. A third option is that it could be added to the remit of the Australian and New Zealand Clinical Trials Registry (ANZCTR). An evaluation of the options is below.

---

123 “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 89.

124 Buch, “Politicians Can’t Have It Both Ways When It Comes to Evidence.”

125 Buck and McGee, “Why Government Needs More Randomised Controlled Trials: Refuting the Myths,” July 2015.

126 “Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings,” 88.

127 Ibid.

128 Ibid., 90–91.

Host	Pros	Cons
<b>ANZCTR</b>	ANZCTR is already setup as a credible trial registry. Standards, procedures and funding streams have already been established.	There may be resistance to including social policy trials on a clinical registry.
<b>Registry within a central agency</b>	Government may be more comfortable keeping the registry within the central team responsible for expanding the use of RCTs. A key advantage of the early form of BIT was described to us as the trust from civil servants that comes with being a civil servant.	There may be less trust in the registry if it is housed centrally, as it may be subject to interference.
<b>Registry hosted outside of government</b>	There is likely to be a high degree of public trust in the registry if it is hosted externally. Non-government social policy trials can also be registered.	Government officials may be dissuaded from running trials, and ongoing funding and maintenance would be concerns. Some particularly sensitive departments would be more likely to welcome a registry if it's safe and internal, but not if it's public.

Table 8 – Assessment of options for hosting a policy RCT registry

We think the government should investigate housing the social policy registry through a university. An externally-hosted registry is the most transparent option, does not require changing an existing, high-functioning clinical registry and is likely to have the requisite expertise.

### 3.3.2. Link Administrative data

#### **Recommendations**

Senior public servants should link data within departments, between departments and across jurisdictions.

As described above, data is the backbone of experimentation, and expensive, novel data has been a significant cost driver of expensive RCTs in the past. To the extent that those looking to run new RCTs should start by looking at existing administrative datasets, those working on building government RCT capability should improve and link existing datasets. The scale outlined in *Figure 6* above describes how linked administrative data provide an opportunity for measuring outcomes that doesn't otherwise exist, is of high quality, and is cheaper than creating new data. Accordingly, linked administrative data is key to expanding the use of RCTs in Australia. The case study in *Figure 10* below from the US shows the potential insight to be gained by combining RCTs with administrative datasets.<sup>129</sup>

129 Chetty, Hendren, and Katz, "The Effects of Exposure to Better Neighborhoods on Children."

A recent report advocating for more use of RCTs in health care delivery argued that “administrative data offer[s] the potential to do high-quality, low-cost, rapid turnaround RCTs”.<sup>130</sup> In addition to the low cost of using existing administrative data, the report outlines five advantages to using administrative data in RCTs:

- Easier to identify participants;
- Many existing databases include a “near-census of the relevant individuals”;
- Less likely to be skewed in any direction as they are collected for purposes separate to the study;
- In many cases is more accurate than survey data and has existing quality control procedures; and;
- Can support long-term outcomes analysis.<sup>131</sup>

A senior public servant described to us how government departments have already begun investing in linking data and improving access protocols. This should continue to happen apace.

A key challenge to using administrative data in Australia was highlighted by Cobb-Clark: In order to run an RCT for a preschool programme with outcome measures into adulthood, the measures would cross numerous state datasets, possibly across states, and certainly into federal datasets.<sup>132</sup> As Australia does not have a national identity card and children are not given Tax File Numbers at birth, such data linkage is presently very challenging. Accordingly, so long as the data infrastructure of government administrative datasets does not support the outcomes measurement needed for policy evaluation, data creation (mostly by survey) by researchers and programme managers may be inevitable – and expensive.<sup>133</sup>

To that end, the government should prioritise linking data between departments. This investment can dramatically reduce the cost of evaluating programmes across government.

#### **Case Study - USA**

In the mid-1990s, an experimental program called Moving to Opportunity offered a random selection of families who were living in poverty vouchers to move to more affluent neighborhoods. In 2015, a study combined data from the program with administrative data from tax returns and found significant results. They found that children who had moved house under the age of 13 had higher university attendance rates, lived in better neighborhoods as adults, were less likely to be single parents, and had 31% higher annual incomes on average. The study found that the “gains from moving fall with the age when children move”.

This program is a powerful example of the transformative potential of social programs and the value of administrative data matching.

*Figure 10 - Moving to Opportunity in the US*

130 Finkelstein and Taubman, “Using Randomised Evaluations to Improve the Efficiency of US Healthcare Delivery,” 4.

131 Ibid., 16–17.

132 Cobb-Clark, interview, November 8, 2015.

133 Ibid.

## Conclusion

Australian Federal Government ministers and public servants have an opportunity to improve lives through more effective government. Social programmes across different contexts have shown a 20% reduction in child abuse, an 11% increase in average annual earnings and a 40% reduction in unintended pregnancies.<sup>134</sup> These claims are possible because of the rigorous evidence provided by RCTs.

With the launch of BETA in the federal government, BIU in the NSW government and a newly announced behavioural unit in the Victorian government, there is an opportunity to translate the growth of behavioural insights in governments into growth of RCTs in government. These units have the capacity to act as centres of excellence driving both optimisation and policy RCTs in those governments.

And we think there is more political support for such RCTs than is often supposed. Our survey found that 73% of Australian parliamentarians support the use of RCTs in social policy. Further, of those who had used RCTs to inform their policy decisions in the past, 67% chose RCTs as one of their top three priority policy inputs for the future. This suggests that demand for RCTs may grow as experience with and exposure to RCTs grow.

To take this opportunity, start now. While there is work to be done building long-term government capability for RCTs, we recommend building that capability while developing experience and familiarity with RCTs. Senior public servants should start by commissioning small, uncontroversial, high impact/cost likelihood, service optimisation RCTs. Ministers should start pursuing an RCT in social policy. Start with RCTs that are likely to be successful, and start in areas likely to provide value for money for government. Interventions which have their outcome measures tracked by administrative data, which have large target populations and which are easily standardised are ideal.

To make RCTs a routine part of government, Ministers should demand evidence and RCT evaluation strategies from public servants and allocate funding based on evidence. Senior public servants should push available evidence to ministers and outline the most rigorous possible evaluation strategies for new policies. BETA should become a 'centre of excellence' for RCTs and facilitate RCTs across Federal Government departments and agencies. Trials should be pre-registered and results should be published, even if they show policies had no effect.

Establishing a rigorous, evidence-based policy making culture in Australia requires courage, as many RCTs will show no effect. It requires humility, as policy makers must admit that they do not have all the answers. And it requires leadership, if politicians and public servants are to turn the promise of RCTs into reality. Effective social policy can save and improve lives and deliver a fairer and more equal society. It is too important to not know what works.

---

*"If we do our job well, there will be governments where RCTs will be the normal way of doing things. That is, any time they're debating what letter to use or email, they'll think of an RCT as their way of answering that question."*<sup>135</sup>

---

---

<sup>134</sup> Nurse-Family partnership, Career Academies and Carrera Adolescent Pregnancy Prevention Program from "Top Tier Evidence", <http://toptierevidence.org/>, accessed 28 February 2016.

<sup>135</sup> Linos, interview.

## 4. Appendix

### 4.1. Survey of Australian Parliamentary Views of RCTs

	Australia	UK
<i>How much do you support or oppose the following?</i>	Percent overall	<b>support</b>
<b>Support RCTs:</b> <i>The use of controlled experiments or trials to design and test more areas of government social policy</i>	73%	67%
<b>Support pilots:</b> <i>The use of pilot schemes without control groups to design and test more areas of government social policy</i>	49%	64%
<i>How much do you agree or disagree with the following statements?</i>	Percent overall	<b>disagree</b>
<b>Fairness:</b> <i>Randomly choosing whether some people get a policy intervention and others do not is unfair</i>	36%	41%
<b>Cost:</b> <i>Controlled experiments or trials are too expensive as ways of designing and testing social policies</i>	62%*	46%*
<b>Expectations:</b> <i>We are going to see much greater use of controlled experiments or trials in designing and testing social policies in the next few years</i>	28%	22%
<b>Priority inputs:</b> <i>Which two or three, if any, of the following should politicians pay most attention to when deciding what should be done?</i>	Percent	<b>prioritised</b>
• Views of experts (e.g. academics or think tanks)	60%	50%
• Findings from controlled experiments or trials	51%***	23%***
• Views of constituents	48%	45%
• Views of practitioners (e.g. teachers, police etc)	35%	40%
• What works in other countries	25%	14%
• Survey research with those affected	26%	38%
• Their own principles	22%	34%
• Findings from pilot schemes without control groups	12%**	31%**
• Their personal experiences	9%	8%
• Views of journalists	0%	0%
<b>Past experience:</b> <i>Which, if any, of the following have you ever used to justify a policy that you support?</i>	Percent	<b>have used</b>
• Views of experts (e.g. academics or think tanks)	87%*	73%*
• Findings from controlled experiments or trials	77%	65%
• Views of constituents	76%	66%
• Views of practitioners (e.g. teachers, police etc)	74%	75%
• What works in other countries	66%	73%
• Survey research with those affected	66%	70%
• Their own principles	58%	43%
• Findings from pilot schemes without control groups	42%*	25%*
• Their personal experiences	39%	35%
• Views of journalists	5%	8%

Table 9 – Summary of survey results<sup>136</sup>

Asterisks represent statistical significance of Australia/UK difference. \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.05  
 Figures have been weighted to adjust for party and jurisdiction in both countries.

136 Survey introduction: "As you know, there are many ways of testing the effectiveness of social policies, in areas such as education, crime, health and welfare. For example, in a controlled experiment or trial, some people are randomly chosen to get a policy intervention and others do not get it at all. These groups are then compared to see the effect the policy has had. On the other hand, a pilot scheme is when a policy is tested with part of a population before being rolled out to the whole target population."

## 4.2. Table of minimum sample sizes

*Minimum sample sizes required to statistically significantly detect differences between intervention and control groups:*

Control Group proportion Intervention Group proportion	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%
10%	1,438	2,270	328	140	80	52	36	24	18	12
20%	248	2,996	3,620	456	178	94	58	38	26	16
30%	116	398	2,070	4,556	536	198	100	58	36	24
40%	68	160	502	4,868	5,076	572	202	98	56	32
50%	46	88	190	560	5,180	5,180	560	190	88	46
60%	32	56	98	202	572	5,076	4,868	502	160	68
70%	24	36	58	100	198	536	4,556	2,070	398	116
80%	16	26	38	58	94	178	456	3,620	2,996	248
90%	12	18	24	36	52	80	140	328	2,270	1,438

*Zooming in to the toughest differences to detect:*

Control Group proportion Intervention Group proportion	45%	47%	49%	51%	53%	55%
46%	128,890	129,306	14,398	5,188	2,646	1,600
48%	14,362	129,618	129,826	14,432	5,192	2,644
50%	5,180	14,420	129,930	129,930	14,420	5,180
52%	2,644	5,192	14,432	129,826	129,618	14,362
54%	1,600	2,646	5,188	14,398	129,306	128,890

*Table 10 – Minimum sample sizes required to detect differences*

‘Proportion’ here refers to a binary measure of interest. For example: Percent of students reading at or above level, or percent of citizens who owe tax who pay within 30 days. If the control group proportion is 75% of students reading at level, and the intervention group proportion is 80% of students reading at level, the ‘effect size’ is 5 percentage points, and will require 3,620 students to be detected (1,810 in each group).<sup>137</sup>

<sup>137</sup> Calculated using G\*Power, using a two-tail z-test for differences between two independent proportions, assuming: Type I ( $\alpha$ ) error significance level of 5%, Type II error ( $\beta$ ) of 5% (power of 95%), intervention and control groups the same size.



## 4.3. Australian RCTs in Social Policy

### *Australian Social Policy RCTs - Not Health Related*

Year	RCT	Policy	Partner Organisation
1987	Milne & Spence, "Training Social Perception Skills with Primary School Children: A Cautionary Note" <sup>138</sup>	Education	(no partner organisation)
1993	Schultz et al., "Psychoeducational Support for Parents of Children with Intellectual Disability: An Outcome Study" <sup>139</sup>	Parenting, Disability	(no partner organisation)
1997	Connell et al., "Self-Directed Behavioral Family Intervention for Parents of Oppositional Children in Rural and Remote Areas" <sup>140</sup>	Parenting	(no partner organisation)
2004	Cross et al., "Australia: The Friendly Schools project" <sup>141</sup>	Education	(no partner organisation)
2003	Breunig et al., "Assisting The Long-Term Unemployed: Results From A Randomised Trial" <sup>142</sup>	Welfare	Federal Department of Family and Community Services
2004	Shanahan et al., "Cost-Effectiveness Analysis Of The New South Wales Adult Drug Court Program" <sup>143</sup>	Crime	NSW Government (various)
2004	Dockery & Stromback, "An Evaluation Of A Parenting Payment Intervention Program" <sup>144</sup>	Parenting	Centrelink, Fed. Dep. of Family & Community Services
2006	Morawska & Sanders, "Self-administered behavioral family intervention for parents of toddlers: Part I" <sup>145</sup>	Parenting	(no partner organisation)
2007	Hunt, "The Effect of an Education Program on Attitudes and Beliefs about Bullying and Bullying Behaviour in Junior Secondary School Students" <sup>146</sup>	Education	Sydney Catholic Education Office
2008	Hiscock et al., "Universal Parenting Programme to Prevent Early Childhood Behavioural Problems: Cluster Randomised Trial" <sup>147</sup>	Parenting	(no partner organisation)
2011	Taft et al., "Mothers' AdvocateS In the Community (MOSAIC) - non-professional mentor support to reduce intimate partner violence and depression in mothers: a cluster randomised trial in primary care." <sup>148</sup>	Domestic Violence	Women's Health West, Berry Street, Domestic Violence Resource Centre Victoria, and the MOSAIC Reference Group
2012	Johnson et al., "Meeting The Challenge? Transitions Out Of Long-Term Homelessness. A Randomised Controlled Trial Examining The 24-Month Costs, Benefits And Social	Homelessness	Sacred Heart Mission (NGO)

138 Milne and Spence, "Training Social Perception Skills with Primary School Children."

139 Schultz et al., "Psychoeducational Support for Parents of Children with Intellectual Disability."

140 Connell, Sanders, and Markie-Dadds, "Self-Directed Behavioral Family Intervention for Parents of Oppositional Children in Rural and Remote Areas."

141 Smith, Pepler, and Rigby, *Bullying in Schools*.

142 Breunig et al., "Assisting the Long-Term Unemployed."

143 Shanahan et al., "Cost-Effectiveness Analysis of the New South Wales Adult Drug Court Program."

144 Dockery and Stromback, "An Evaluation of a Parenting Payment Intervention Program."

145 Morawska and Sanders, "Self-Administered Behavioral Family Intervention for Parents of Toddlers."

146 Hunt, "The Effect of an Education Program on Attitudes and Beliefs about Bullying and Bullying Behaviour in Junior Secondary School Students."

147 Hiscock et al., "Universal Parenting Programme to Prevent Early Childhood Behavioural Problems."

148 Taft et al., "Mothers' AdvocateS In the Community (MOSAIC)- Non-Professional Mentor Support to Reduce Intimate Partner Violence and Depression in Mothers."

	Outcomes From The Journey To Social Inclusion Pilot Program” <sup>149</sup>		
2015	Mazerolle et al., “Optimising The Length Of Random Breath Tests: Results From The Queensland Community Engagement Trial” <sup>150</sup>	Crime	Queensland Police Service
2015	Hunter et al., “Program Fidelity Measures Associated With An Effective Child Restraint Program: Buckle-Up Safely” <sup>151</sup>	Transport	Roads and Traffic Authority (NSW)
RISE (Reintegrative Shaming Experiments)			
1998	Sherman et al., “Experiments in restorative policing: A progress report on the Canberra Reintegrative Shaming Experiments”	Police, Justice	Australian Federal Police
2011	Strang et al., “Experiments in restorative policing: Final report - Canberra Reintegrative Shaming Experiments (RISE)” <sup>152</sup>		
Triple P Parenting Program			
2000	Sanders et al., “The Triple P-Positive Parenting Program: A comparison of enhanced, standard, and self-directed behavioural family intervention for parents of children with early onset conduct problems.” <sup>153</sup>	Parenting, Family	(no partner organisation)
2000	Sanders et al., “The Mass Media and the Prevention of Child Behavior Problems: The Evaluation of a Television Series to Promote Positive Outcomes for Parents and Their Children.” <sup>154</sup>		(no partner organisation)
2003	Martin et al., “Balancing Work and Family: A Controlled Evaluation of the Triple P- Positive Parenting Program as a Work-Site Intervention” <sup>155</sup>		University of Queensland
2005	Gallart et al., “The Effectiveness of Group Triple P and the Impact of the Four Telephone Contacts” <sup>156</sup>		South West Sydney Area Health Service
2006	Markie-Dadds & Sanders, “Self-Directed Triple P (Positive Parenting Program) for Mothers with Children at-Risk of Developing Conduct Problems” <sup>157</sup>		(no partner organisation)
2007	Turner et al., “Randomised clinical trial of a group parent education programme for Australian Indigenous families” <sup>158</sup>		Participating Child Health Centres and Indigenous Heath Services
2009	Joachim et al., “Reducing Preschoolers’ Disruptive Behaviour in Public with a Brief Parent Discussion Group” <sup>159</sup>		(no partner organisation)
2009	Morawska et al., “An evaluation of a behavioural parenting intervention for parents of gifted children” <sup>160</sup>		(no partner organisation)

Table 11 – List of social policies RCTs in Australia – not health related

149 Johnson et al., "Meeting the Challenge? Transitions out of Long-Term Homelessness. A Randomised Controlled Trial Examining the 24-Month Costs, Benefits and Social Outcomes from the Journey to Social Inclusion Pilot Program."

150 Mazerolle et al., "Optimising the Length of Random Breath Tests."

151 Hunter, Brown, and Stevenson, "Program Fidelity Measures Associated With an Effective Child Restraint Program."

152 Strang et al., "Australian Institute of Criminology - Experiments in Restorative Policing: Final Report."

153 Sanders et al., "The Triple P-Positive Parenting Program."

154 Sanders, Montgomery, and Brechman-Toussaint, "The Mass Media and the Prevention of Child Behavior Problems."

155 Martin and Sanders, "Balancing Work and Family."

156 Gallart and Matthey, "The Effectiveness of Group Triple P and the Impact of the Four Telephone Contacts."

157 Markie-Dadds and Sanders, "Self-Directed Triple P (Positive Parenting Program) for Mothers with Children at-Risk of Developing Conduct Problems."

158 Turner, Richards, and Sanders, "Randomised Clinical Trial of a Group Parent Education Programme for Australian Indigenous Families."

159 Joachim, Sanders, and Turner, "Reducing Preschoolers' Disruptive Behavior in Public with a Brief Parent Discussion Group."

160 Morawska and Sanders, "An Evaluation of a Behavioural Parenting Intervention for Parents of Gifted Children."



***Australian Social Policy RCTs - Health Related (excluding clinical trials)***

<b>Year</b>	<b>RCT</b>	<b>Policy Areas</b>
1976	Cullen, "A six-year controlled trial of prevention of children's behavior disorders" <sup>161</sup>	Public Health
1980	Coyne et al., "Evaluation of preschool meals programmes on the nutritional health of Aboriginal children" <sup>162</sup>	Public Health Indigenous Affairs
1988	Marsh & Peart, "Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and multidimensional self-concepts" <sup>163</sup>	Public Health
1989	King et al., "School Refusal - Graduated And Rapid Behavioral Treatment Strategies" <sup>164</sup>	Education Health
1994	Richardson et al., "Participation In Breast Cancer Screening: Randomised Controlled Trials Of Doctors' Letters And Of Telephone Reminders" <sup>165</sup>	Health
1997	Schofield et al., "Interventions With Retailers To Reduce Cigarette Sales To Minors: A Randomised Controlled Trial" <sup>166</sup>	Public Health
1997	Greaves, "The effect of rational-emotive parent education on the stress of mothers of children with Down Syndrome" <sup>167</sup>	Family Health
2000	Board et al., "A Randomised Controlled Trial Of The Costs Of Hospital As Compared With Hospital In The Home For Acute Medical Patients" <sup>168</sup>	Health
2000	Mitchell et al., "A Randomised Trial Of An Intervention To Develop Health Promoting Schools In Australia: The South Western Sydney Study" <sup>169</sup>	Public Health Education
2000	Peel et al., "Home Safety Assessment In The Prevention Of Falls Among Older People" <sup>170</sup>	Public Health
2000	Fraser et al., "Home Visiting Intervention For Vulnerable Families With Newborns: Follow-Up Results Of A Randomised Controlled Trial" <sup>171</sup>	Welfare Public Health
2002	Heyne et al., "Evaluation of Child Therapy and Caregiver Training in the Treatment of School Refusal" <sup>172</sup>	Education Health
2003	Dolan, "A Randomised Controlled Trial Of Methadone Maintenance Treatment In Nsw Prisons" <sup>173</sup>	Crime Health
2003	Llewellyn et al., "Promoting health and home safety for children of parents with intellectual disability: a randomised controlled trial" <sup>174</sup>	Education, Family Public Health
2004	Bogossian, "The Mothers' Health Study : A Randomised Controlled Trial Of A Social Support Intervention On The Health Of Mothers In The Year After Birth" <sup>175</sup>	Health

161 Cullen, "A Six-Year Controlled Trial of Prevention of Children's Behavior Disorders."

162 Coyne, Dowling, and Condon-Paoloni, "Evaluation of Preschool Meals Programmes on the Nutritional Health of Aboriginal Children."

163 Marsh and Peart, "Competitive and Cooperative Physical Fitness Training Programs for Girls."

164 King and Ollendick, "School Refusal - Graduated and Rapid Behavioral Treatment Strategies."

165 Richardson et al., "Participation in Breast Cancer Screening."

166 Schofield, Sanson-Fisher, and Gulliver, "Interventions with Retailers to Reduce Cigarette Sales to Minors."

167 Greaves, "The Effect of Rational-Emotive Parent Education on the Stress of Mothers of Young Children with Down Syndrome."

168 Board, Brennan, and Caplan, "A Randomised Controlled Trial of the Costs of Hospital as Compared with Hospital in the Home for Acute Medical Patients."

169 Mitchell et al., "A Randomised Trial of an Intervention to Develop Health Promoting Schools in Australia."

170 Peel, Steinberg, and Williams, "Home Safety Assessment in the Prevention of Falls among Older People."

171 Fraser et al., "Home Visiting Intervention for Vulnerable Families with Newborns."

172 HEYNE et al., "Evaluation of Child Therapy and Caregiver Training in the Treatment of School Refusal."

173 Kate Dolan, "A Randomised Controlled Trial of Methadone Maintenance Treatment in NSW Prisons."

174 Llewellyn et al., "Promoting Health and Home Safety for Children of Parents with Intellectual Disability."

175 Bogossian, "The Mothers' Health Study."

Year	RCT	Policy Areas
2005	Calver et al., "Does Health Assessment Improve Health Outcomes In Indigenous People? An Rct With 13 Years Of Follow-Up" <sup>176</sup>	Health Indigenous Affairs
2006	Digiusto et al., "Effects Of Pharmacotherapies For Opioid Dependence On Participants' Criminal Behaviour And Expenditure On Illicit Drugs: An Australian National Evaluation (Nepod)" <sup>177</sup>	Crime Health
2007	Ireland et al., "Efficacy of Written Emotional Expression in the Reduction of Psychological Distress in Police Officers" <sup>178</sup>	Police Health
2008	Wen et al., "Increasing Active Travel To School: Are We On The Right Track? A Cluster Randomised Controlled Trial From Sydney, Australia" <sup>179</sup>	Public Health Education
2009	Nagel et al., "Approach To Treatment Of Mental Illness And Substance Dependence In Remote Indigenous Communities: Results Of A Mixed Methods Study" <sup>180</sup>	Health Indigenous Affairs
2010	Skouteris et al., "Healthy Eating And Obesity Prevention For Preschoolers: A Randomised Controlled Trial" <sup>181</sup>	Education Public Health
2011, 12, 15	Wen et al., "Sustainability Of Effects Of An Early Childhood Obesity Prevention Trial Over Time: A Further 3-Year Follow-Up Of The Healthy Beginnings Trial," <sup>182, 183, 184</sup>	Public Health
2011	Waters et al., "An Exploratory Cluster Randomised Controlled Trial Of Knowledge Translation Strategies To Support Evidence-Informed Decision-Making In Local Governments (The Kt4lg Study)" <sup>185,186</sup>	Public Health Local Government
2011	Kemp et al., "Child And Family Outcomes Of A Long-Term Nurse Home Visitation Programme: A Randomised Controlled Trial" <sup>187</sup>	Health
2011	Purcell et al., "Is education an effective management strategy for reducing cancer-related fatigue?" <sup>188</sup>	Health
2012	Parker et al., "An Oral Health Literacy Intervention For Indigenous Adults In A Rural Setting In Australia" <sup>189</sup>	Public Health Indigenous Affairs
2013	Burford et al., "Internet-Based Photoaging Within Australian Pharmacies to Promote Smoking Cessation: Randomised Controlled Trial" <sup>190</sup>	Public Health
2015	Taft et al., "Maternal And Child Health Nurse Screening And Care For Mothers Experiencing Domestic Violence (Move): A Cluster Randomised Trial"	Health Domestic Violence

*Table 12 – List of social policies RCTs in Australia – health related*

176 Calver et al., "Does Health Assessment Improve Health Outcomes in Indigenous People?"

177 Digiusto et al., "Effects of Pharmacotherapies for Opioid Dependence on Participants' Criminal Behaviour and Expenditure on Illicit Drugs."

178 Ireland, Malouff, and Byrne, "The Efficacy of Written Emotional Expression in the Reduction of Psychological Distress in Police Officers."

179 Wen et al., "Increasing Active Travel to School."

180 Nagel et al., "Approach to Treatment of Mental Illness and Substance Dependence in Remote Indigenous Communities."

181 Skouteris et al., "Healthy Eating and Obesity Prevention for Preschoolers."

182 Wen et al., "Effectiveness of an Early Intervention on Infant Feeding Practices and 'Tummy Time.'"

183 Wen et al., "Effectiveness of Home Based Early Intervention on Children's BMI at Age 2."

184 Wen et al., "Sustainability of Effects of an Early Childhood Obesity Prevention Trial Over Time."

185 Waters et al., "An Exploratory Cluster Randomised Controlled Trial of Knowledge Translation Strategies to Support Evidence-Informed Decision-Making in Local Governments (The KT4LG Study)."

186 Armstrong et al., "Knowledge Translation Strategies to Improve the Use of Evidence in Public Health Decision Making in Local Government."

187 Kemp et al., "Child and Family Outcomes of a Long-Term Nurse Home Visitation Programme."

188 Purcell et al., "Is Education an Effective Management Strategy for Reducing Cancer-Related Fatigue?"

189 Parker et al., "An Oral Health Literacy Intervention for Indigenous Adults in a Rural Setting in Australia."

190 Burford et al., "Internet-Based Photoaging Within Australian Pharmacies to Promote Smoking Cessation."

## 4.4. Successful International Social Policy RCT Examples

Trial	Key Findings
<b>Career Academies</b> <sup>191</sup>	<p><b>Intervention:</b> “Career Academies are organized as small learning communities, combine academic and technical curricula around a career theme, and establish partnerships with local employers to provide work-based learning opportunities.”</p> <p><b>Key Findings:</b> “Produced sustained earnings gains that averaged 11% (or \$2,088) more per year for Academy group members than for individuals in the non-Academy group — a \$16,704 boost in total earnings over the eight years of follow-up (in 2006 dollars).”</p>
<b>Carrera Adolescent Pregnancy Prevention Program</b> <sup>192</sup>	<p><b>Intervention:</b> “A year-round after-school program with a comprehensive youth development orientation.”</p> <p><b>Key Findings:</b> “Female program participants had significantly lower odds than controls of being sexually-active [50% lower] and of having experienced a pregnancy [30% lower].”</p>
<b>H&amp;R Block College Financial Aid Application Assistance</b> <sup>193</sup>	<p><b>Intervention:</b> “Low-income individuals receiving tax preparation help were also offered immediate assistance and a streamlined process to complete the Free Application for Federal Student Aid (FAFSA) for themselves or their children.”</p> <p><b>Key Findings:</b> “Students just graduating from high school whose parents received the assistance experienced an 8 percentage point increase in college enrollment the following year.”</p>
<b>Treatment Foster Care Oregon</b> <sup>194</sup>	<p><b>Intervention:</b> Multidimensional Treatment Foster Care (MTFC)</p> <p><b>Key Findings:</b> “Fewer post-baseline pregnancies were reported for MTFC girls (26.9%) than for GC girls (46.9%)—an effect that remained significant after controlling for baseline criminal referrals, pregnancy history, and sexual activity.”</p>
<b>Nurse-Family Partnership</b> <sup>195</sup>	<p><b>Intervention:</b> “Prenatal and infancy/toddler home visits by paraprofessionals and by nurses on child development at child ages 6 and 9 years.”</p> <p><b>Key Findings:</b> “Nurse-visited children were [45% as] likely to be classified as having total emotional/behavioral problems at age 6 years, [44% as likely to be] internalizing problems at age 9 years, and [34% as likely to have] dysfunctional attention at age 9 years.”</p>
<b>Staying Free</b> <sup>196</sup>	<p><b>Intervention:</b> “Advice from physicians and nurses and 2 pamphlets, [as well as] 60 minutes of bedside counselling, take-home materials and 7 nurse-initiated counselling calls for 2 months after discharge.”</p> <p><b>Key Findings:</b> “Continuous 12-month abstinence was 57% in the intensive group and 39% in the [control] group (p &lt; 0.01).”</p>
<b>Success for All for grades K-2</b> <sup>197</sup>	<p><b>Intervention:</b> “Schoolwide intervention that focuses on prevention and early, intensive intervention designed to detect and resolve reading problems as early as possible”</p> <p><b>Key Findings:</b> “The program increased second-grade reading achievement in Success for All schools by 25-30% of a grade-level, three years after random assignment.”</p>

Table 13 – List of select RCTs in social policy, curated from website Top Tier Evidence<sup>198</sup>

191 Kemple, “Career Academies.”

192 Philliber S et al., “Preventing Pregnancy and Improving Health Care Access among Teenagers.”

193 Bettinger et al., “The Role of Application Assistance and Information in College Decisions.”

194 Kerr, Leve, and Chamberlain, “Pregnancy Rates among Juvenile Justice Girls in Two Randomised Controlled Trials of Multidimensional Treatment Foster Care.”

195 Olds DL et al., “Effects of Home Visits by Paraprofessionals and by Nurses on Children.”

196 Smith and Burgess, “Smoking Cessation Initiated during Hospital Stay for Patients with Coronary Artery Disease.”

197 Borman et al., “Final Reading Outcomes of the National Randomised Field Trial of Success for All.”

198 “Top Tier Evidence.”

## 4.5. Research Appendices

### 3.5.1. Methodology

Our findings and recommendations have been formed as a result of extensive research, both qualitative and quantitative, summarised below with additional details throughout this section.

#### Qualitative research:

- *Interviews and case studies*

We have interviewed leading practitioners, academics and private sector professionals to understand the common factors leading to successful implementation of RCTs. We distil insights across these cases, both of successes and failures. The identification of threads and common lessons from these interviews forms the bulk of our research. See the interview list in the appendix for more details.

- *Literature review*

We have reviewed a wide range of sources globally and from Australia focussed on both government and non-government sectors. These include approaches to running and spreading trials in government, the social and private sector. They also include implementations of social impact bonds, incorporation of behavioural insights, transitions to digital government, leading change in large organisations, lean start-up and marketing methodologies.

#### Quantitative research:

- *Survey of Australian and British parliamentarians*

Ipsos MORI completed an in-person survey of UK parliamentarians to assess their understanding and attitudes towards RCTs and experimentation. Based on that survey we ran an online survey of Australia's parliamentarians across the federal, state and territory governments, asking for views relating to randomised controlled trials (RCTs) in social policy. The survey attracted 109 responses from 828 invitations (13% response rate). The results have been weighted to adjust for representation by party and jurisdiction.

- *Education Endowment Fund (EEF) (UK) Project Database<sup>199</sup>*

The EEF is one of the UK National Government's What Works Centres. The EEF both consolidates education related research with a public registry of findings, and also directly commissions RCTs. The EEF listed 119 projects, most of which were RCTs, at the start of 2016.

- *American Economic Association Social Science Registry<sup>200</sup>*

Since 2012, the American Economic Association has hosted a public social science registry for RCTs, primarily used by academics and practitioners in the USA. At the beginning of 2016 there were 594 studies in 88 countries.

- *Campbell Collaboration Systematic Reviews<sup>201</sup>*

Since 2004, this site publishes systematic reviews, primarily of RCTs in Crime & Justice, Education, International Development, and Social Welfare. This site was the anchor for determining the relative RCT production of different countries around the world, as the reviews are based on comprehensive global searches for high quality RCT by policy domain.

---

199 "About."

200 "AEA RCT Registry."

201 "The Campbell Collaboration."

### 3.5.2. Interview List

- **Jon Baron**, Vice-President of Evidence-Based Policy at the Arnold Foundation, formerly the founder and president of the Coalition for Evidence-Based Policy
- **Max Bazerman**, Jesse Isidor Straus Professor of Business Administration at the Harvard Business School and the Co-Director of the Center for Public Leadership at the Harvard Kennedy School
- **Iris Bohnet**, Professor of Public Policy at Harvard Kennedy School, Director of Women and Public Policy Program at Harvard Kennedy School, Co-chair of the Behavioral Insights Group and Associate Director of the Harvard Decision Science Laboratory
- **Martin Bowles**, Secretary, Department of Health of the Australian Government
- **Kieron Boyle**, Head of Social Investment and Finance at the UK Cabinet Office
- **Prateek Buch**, Public Policy Associate, Sense About Science UK
- **Deborah Cobb-Clark**, Professor of Economics at the University of Sydney, former Director of the Melbourne Institute of Applied Economic and Social Research
- **Abigail Dalton**, Assistant Director, Behavioral Insights Group, Center for Public Leadership at Harvard Kennedy School of Government
- **Jorrit de Jong**, Lecturer in Public Policy and Academic Director, Innovations in Government Program, Ash Center for Democratic Governance and Innovation, Harvard Kennedy School of Government
- **Andrey Fradkin**, Part-time data scientist at Airbnb and Postdoctoral associate at the MIT Initiative on the Digital Economy
- **Rory Gallagher**, Managing Director, Behavioural Insights Team Australia and Asia-Pacific
- **Gloria Gong**, Assistant Director, Government Performance Lab, Taubman Center for State and Local Government, Harvard Kennedy School of Government
- **Michael Hiscox**, Director, Behavioural Economics Team of Australia
- **Andrew Leigh**, Federal Member for Fraser in the Parliament of Australia, Shadow Assistant Treasurer and Shadow Minister for Competition in the Australian Labor Party, Ph.D. and M.P.A. from Harvard Kennedy School, Fellow of the Australian Academy of Social Sciences
- **Elizabeth Linos**, Principal Advisor, Head of Research and Evaluation at the Behavioural Insights Team North America
- **Mike Luca**, Assistant Professor of Business Administration at Harvard Business School, Faculty Affiliate, Behavioral Insights Group, Harvard Kennedy School and Faculty Affiliate, Ideas42
- **Jason McNamara**, Executive Director, Office of Best Practice Regulation, Department of Prime Minister and Cabinet of the Australian Government
- **Ted Robertson**, Managing Director, Ideas42 working strategy and application of behavioral science to health care, city government, and national civics
- **Todd Rogers**, Associate Professor of Public Policy, Center for Public Leadership, Harvard Kennedy School of Government and Director of the Student Social Support R&D Lab
- **Michael Sanders**, Principal Advisor & Head of Research at the Behavioural Insights Team, UK
- **Kathy Stack**, Advisor for Evidence-Based Innovation at the U.S. Office of Management and Budget, formerly OMB's Deputy Associate Director for Education, Income Maintenance, and Labor
- **Stephanie Wade**, Director, Innovation Lab in the Office of Personnel Management, US Government
- **Mitch Weiss**, Senior Lecturer in the Entrepreneurial Management unit at the Harvard Business School, formerly the Chief of Staff and partner to Boston's former Mayor Thomas Menino
- **Bob Wu**, Advisor, Behavioral Economics Team of the Australian Government
- Senior Public Servant, Department of Premier and Cabinet, NSW Government



### 3.5.3. References

- "About." *Education Endowment Foundation*. Accessed March 5, 2016. <https://educationendowmentfoundation.org.uk/about/>.
- "AEA RCT Registry." Accessed February 11, 2016. <https://www.socialscienceregistry.org/>.
- "Applying Behavioural Insights to Reduce Fraud, Error and Debt." Cabinet Office Behavioural Insights Team, February 2012.
- Armstrong, Rebecca, Elizabeth Waters, Maureen Dobbins, Laurie Anderson, Laurence Moore, Mark Petticrew, Rachel Clark, et al. "Knowledge Translation Strategies to Improve the Use of Evidence in Public Health Decision Making in Local Government: Intervention Design and Implementation Plan." *Implementation Science* 8 (2013): 121. doi:10.1186/1748-5908-8-121.
- Ayotte, Kelly, Mark Warner, Glenn Hubbard, Gene Sperling, Melody Barnes, John Bridgeland, Kevin Madden, and Howard Wolfson. *Moneyball for Government*. Edited by Jim Nussle and Peter Orszag. Disruption Books, 2014.
- Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein. "The Oregon Experiment — Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* 368, no. 18 (May 2, 2013): 1713–22. doi:10.1056/NEJMsa1212321.
- Barber, Michael. *How to Run a Government: So That Citizens Benefit and Taxpayers Don't Go Crazy*. UK: Penguin UK, 2015.
- Baron, Jon. "Board of Advisors of the Coalition for Evidence-Based Policy," April 24, 2015.
- "Better Indigenous Policies: The Role of Evaluation - Roundtable Proceedings." Productivity Commission, October 23, 2012.
- Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment\*." *The Quarterly Journal of Economics* 127, no. 3 (August 1, 2012): 1205–42. doi:10.1093/qje/qjs017.
- Board, Neville, Nicholas Brennan, and Gideon A. Caplan. "A Randomised Controlled Trial of the Costs of Hospital as Compared with Hospital in the Home for Acute Medical Patients." *Australian and New Zealand Journal of Public Health* 24, no. 3 (June 2000): 305–11.
- Bogossian, Fiona E. "The Mothers' Health Study : A Randomised Controlled Trial of a Social Support Intervention on the Health of Mothers in the Year after Birth," 2004. <http://espace.library.uq.edu.au/view/UQ:254235>.
- Borman, Geoffrey D., Robert E. Slavin, Alan C. K. Cheung, Anne M. Chamberlain, and et al. "Final Reading Outcomes of the National Randomized Field Trial of Success for All." *American Educational Research Journal* 44, no. 3 (September 2007): 701–31.
- Breunig, Robert, Yvonne Dunlop, Deborah A. Cobb-Clark, and Marion Terrill. "Assisting the Long-Term Unemployed: Results from a Randomised Trial." *Economic Record* 79, no. 244 (March 2003): 84–102.
- Buch, Prateek. "Politicians Can't Have It Both Ways When It Comes to Evidence." *Sense about Science*. Accessed March 11, 2016. <http://www.senseaboutscience.org/blog.php/85/politicians-cant-have-it-both-ways-when-it-comes-to-evidence>.
- Buck, Stuart, and Josh McGee. "Why Government Needs More Randomized Controlled Trials: Refuting the Myths." Laura and John Arnold Foundation, July 2015.
- "Building the Evidence Base for What Works." *Whitehouse.gov*, December 17, 2013. <https://www.whitehouse.gov/blog/2013/12/17/building-evidence-base-what-works>.
- Burford, Oksana, Moyez Jiwa, Owen Carter, Richard Parsons, and Delia Hendrie. "Internet-Based Photoaging Within Australian Pharmacies to Promote Smoking Cessation: Randomized Controlled Trial." *Journal of Medical Internet Research* 15, no. 3 (March 26, 2013): e64. doi:10.2196/jmir.2337.
- Burtless, Gary. "Innovations in Labour Market Policies: The Australian Way Comments on an OECD Report." *Australian Economic Review* 35, no. 1 (March 1, 2002): 97–103. doi:10.1111/1467-8462.00227.
- Calver, Janine, Amy Wiltshire, C. D'Arcy J. Holman, and Ernest Hunter et al. "Does Health Assessment Improve Health Outcomes in Indigenous People? An RCT with 13 Years of Follow-Up." *Australian and New Zealand Journal of Public Health* 29, no. 2 (April 2005): 107–11.
- Cherney, Adrian, and Brian Head. "Evidence-Based Policy and Practice Key Challenges for Improvement." *Australian Journal of Social Issues* 45, no. 4 (Summer 2010): 509–26, 439–40.
- Chetty, Raj, Nathaniel Hendren, and Lawrence Katz. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." NBER Working Paper. National Bureau of Economic Research, Inc, May 2015. <http://econpapers.repec.org/paper/nbrnberwo/21156.htm>.
- "Coalition-Board-of-Advisors-Update-04-24-15.pdf," n.d.
- Cobb-Clark, Deborah. "Evidence-Based Policy: The Need for Data and Transparent Evaluations." presented at the German Economic Association Annual Meeting, September 10, 2014.

- "Cognitive Behavioural Therapy (CBT)." Accessed March 1, 2016. <http://whatworks.college.police.uk/toolkit/Pages/Intervention.aspx?InterventionID=32>.
- Connell, Sheryl, Matthew R. Sanders, and Carol Markie-Dadds. "Self-Directed Behavioral Family Intervention for Parents of Oppositional Children in Rural and Remote Areas." *Behavior Modification* 21, no. 4 (October 1, 1997): 379–408. doi:10.1177/01454455970214001.
- Cowen, Tyler. "How Good Are RCTs in Economics? - Marginal Revolution," May 9, 2010. <http://marginalrevolution.com/marginalrevolution/2010/05/how-good-are-rcts.html>.
- Coyne, T., M. Dowling, and D. Condon-Paoloni. "Evaluation of Preschool Meals Programmes on the Nutritional Health of Aboriginal Children." *The Medical Journal of Australia* 2, no. 7 (1980): 369–75.
- Cullen, K. J. "A Six-Year Controlled Trial of Prevention of Children's Behavior Disorders." *The Journal of Pediatrics* 88, no. 4, Part 1 (April 1976): 662–66. doi:10.1016/S0022-3476(76)80032-7.
- Digiusto, Erol, Anthony P. Shakeshaft, Alison Ritter, Richard P. Mattick, Jason White, Nicholas Lintzeris, James Bell, John B. Saunders, and NEPOD Research Group. "Effects of Pharmacotherapies for Opioid Dependence on Participants' Criminal Behaviour and Expenditure on Illicit Drugs: An Australian National Evaluation (NEPOD)." *Australian & New Zealand Journal of Criminology* 39, no. 2 (August 1, 2006): 171–89. doi:10.1375/acri.39.2.171.
- Dockery, A. M., and Thorsten Stromback. "An Evaluation of a Parenting Payment Intervention Program." *Australian Journal of Social Issues* 39, no. 4 (November 2004): 431–42.
- Duflo, Annie, and Dean Karlan. "What Data Can Do to Fight Poverty." *The New York Times*, January 29, 2016.
- "Evaluation Glossary." *Education Endowment Foundation*. Accessed March 5, 2016. <https://educationendowmentfoundation.org.uk/evaluation/the-eefs-approach-to-evaluation/evaluation-glossary/>.
- "Executive Order -- Using Behavioral Science Insights to Better Serve the American People." *Whitehouse.gov*, September 15, 2015. <https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american>.
- "Factiva.com." Accessed March 7, 2016. <https://global.factiva.com/factiva/login/login.asp?productname=global>.
- "Field\_experiments\_in\_rural\_finance.pdf," n.d.
- Finkelstein, Amy, and Sarah Taubman. "Using Randomized Evaluations to Improve the Efficiency of US Healthcare Delivery," February 2015.
- Fraser, Jennifer A, Kenneth L Armstrong, Jeanette P Morris, and Mark R Dadds. "Home Visiting Intervention for Vulnerable Families with Newborns: Follow-up Results of a Randomized Controlled Trial." *Child Abuse & Neglect* 24, no. 11 (November 2000): 1399–1429. doi:10.1016/S0145-2134(00)00193-9.
- "Frequently Asked Questions." *KIPP Public Charter Schools, Knowledge Is Power Program*. Accessed February 11, 2016. <http://www.kipp.org/about-kipp/faq>.
- Gallart, Sonia C., and Stephen Matthey. "The Effectiveness of Group Triple P and the Impact of the Four Telephone Contacts." *Behaviour Change* 22, no. 2 (2005): 71–80.
- Gerber, Alan S., and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. 1st ed. New York: W. W. Norton, 2012.
- Gertler, Paul. "The Impact of Progres on Health; Final Report." In *International Food Policy Research Institute*. AgEcon Search, 2000.
- Glennerster, Rachel. "So You Want to Do an RCT with a Government: Things You Should Know — Running Randomized Evaluations: A Practical Guide," December 9, 2015. <http://runningres.com/blog/2015/12/9/so-you-want-to-do-an-rct-with-a-government-things-you-should-know>.
- Glennerster, Rachel, and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press, 2013.
- Greaves, Daryl. "The Effect of Rational-Emotive Parent Education on the Stress of Mothers of Young Children with Down Syndrome." *Journal of Rational-Emotive and Cognitive-Behavior Therapy* 15, no. 4 (December 1997): 249–67. doi:10.1023/A:1025063425172.
- Gruen, Nicholas. "Beyond Central Planning: Innovation in Government in the 21st Century." *Australian Economic Review* 42, no. 1 (March 1, 2009): 96–103. doi:10.1111/j.1467-8462.2009.00539.x.
- Halpern, David. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. Ebury Press, 2015.
- Harper, Hugo. "Applying Behavioural Insights to Organ Donation: Preliminary Results from a Randomised Controlled Trial." Behavioural Insights Team, Cabinet Office, UK, 2013.
- . *Show Me the Evidence: Obama's Fight for Rigor and Evidence in Social Policy*. Washington, D.C: Brookings Institution Press, 2015.
- Hauser, Oliver, and Michael Luca. "How to Design (and Analyze) a Business Experiment." *Harvard Business Review*, October 29, 2015.



- Hausmann, Ricardo. "The Problem With Evidence-Based Policies." *Project Syndicate*, February 25, 2016. <https://www.project-syndicate.org/commentary/evidence-based-policy-problems-by-ricardo-hausmann-2016-02>.
- Haynes, Laura C., Donald P. Green, Rory Gallagher, Peter John, and David J. Torgerson. "Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages." *Journal of Policy Analysis and Management* 32, no. 4 (September 1, 2013): 718–30. doi:10.1002/pam.21717.
- Haynes, Laura, Owain Service, Ben Goldacre, and David Torgerson. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials." UK Cabinet Office, Behavioural Insights Team, 2012.
- Hearst, Norman, Thomas B. Newman, and Stephen B. Hulley. "Delayed Effects of the Military Draft on Mortality." *New England Journal of Medicine* 314, no. 10 (March 6, 1986): 620–24. doi:10.1056/NEJM198603063141005.
- Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." National Bureau of Economic Research, 2015. <http://www.nber.org/papers/w21178>.
- Heyne, David, Neville J. King, Bruce J. Tonge, Stephanie Rollings, Dawn Young, Melinda Pritchard, and Thomas H. Ollendick. "Evaluation of Child Therapy and Caregiver Training in the Treatment of School Refusal." *Journal of the American Academy of Child & Adolescent Psychiatry* 41, no. 6 (June 2002): 687–95. doi:10.1097/00004583-200206000-00008.
- Hiscock, Harriet, Jordana K. Bayer, Anna Price, Obioha C. Ukoumunne, Susan Rogers, and Melissa Wake. "Universal Parenting Programme to Prevent Early Childhood Behavioural Problems: Cluster Randomised Trial." *BMJ: British Medical Journal* 336, no. 7639 (2008): 318–21.
- "How Good Are RCTs in Economics? - Marginal REVOLUTION." Accessed February 9, 2016. <http://marginalrevolution.com/marginalrevolution/2010/05/how-good-are-rcts.html>.
- Hunt, Caroline. "The Effect of an Education Program on Attitudes and Beliefs about Bullying and Bullying Behaviour in Junior Secondary School Students." *Child and Adolescent Mental Health* 12, no. 1 (February 1, 2007): 21–26. doi:10.1111/j.1475-3588.2006.00417.x.
- Hunter, Kate, Julie Brown, and mark stevenson. "Program Fidelity Measures Associated With an Effective Child Restraint Program: Buckle-Up Safely." *American Journal of Public Health*, March 1, 2015, sec. RESEARCH AND PRACTICE.
- "In Praise of Human Guinea Pigs." *The Economist*, December 12, 2015. <http://www.economist.com/news/leaders/21679799-doctors-use-evidence-when-prescribing-treatments-policymakers-should-too-praise-human>.
- Ipsos Mori. "Are MPs Open to Experimenting? Findings from the Winter 2014 MPs Survey." April 13, 2015.
- Ireland, Matt, John Malouff, and Brian Byrne. "The Efficacy of Written Emotional Expression in the Reduction of Psychological Distress in Police Officers." *International Journal of Police Science & Management* 9, no. 4 (2007): 303–11.
- Joachim, Sabine, Matthew R. Sanders, and Karen M. T. Turner. "Reducing Preschoolers' Disruptive Behavior in Public with a Brief Parent Discussion Group." *Child Psychiatry & Human Development* 41, no. 1 (July 25, 2009): 47–60. doi:10.1007/s10578-009-0151-z.
- Johnson, G, D Kuehnle, S Parkinson, and Y Tseng. "Meeting the Challenge? Transitions out of Long-Term Homelessness. A Randomised Controlled Trial Examining the 24-Month Costs, Benefits and Social Outcomes from the Journey to Social Inclusion Pilot Program." Sacred Heart Mission, St Kilda, 2012.
- Kate Dolan, James Shearer. "A Randomized Controlled Trial of Methadone Maintenance Treatment in NSW Prisons," 2003.
- Kemple, James. "Career Academies: Long-Term Impacts on Work, Education, and Transitions to Adulthood." Text, June 2008. <http://www.mdrc.org/publication/career-academies-long-term-impacts-work-education-and-transitions-adulthood>.
- Kemp, Lynn, Elizabeth Harris, Catherine McMahon, Stephen Matthey, Graham Vimpani, Teresa Anderson, Virginia Schmied, Henna Aslam, and Siggi Zapart. "Child and Family Outcomes of a Long-Term Nurse Home Visitation Programme: A Randomised Controlled Trial." *Archives of Disease in Childhood* 96, no. 6 (June 1, 2011): 533–40. doi:10.1136/adc.2010.196279.
- Kerr, David C. R., Leslie D. Leve, and Patricia Chamberlain. "Pregnancy Rates among Juvenile Justice Girls in Two Randomized Controlled Trials of Multidimensional Treatment Foster Care." *Journal of Consulting and Clinical Psychology* 77, no. 3 (June 2009): 588–93. doi:10.1037/a0015289.
- King, Nj, and Th Ollendick. "School Refusal - Graduated and Rapid Behavioral Treatment Strategies." *Australian And New Zealand Journal Of Psychiatry* 23, no. 2 (1989): 213–23.
- "Laura and John Arnold Foundation Announces Expanded Funding for Low-Cost Randomized Controlled Trials to Drive Effective Social Spending," December 7, 2015. <http://www.arnoldfoundation.org/laura-and-john>

- arnold-foundation-announces-expanded-funding-for-low-cost-randomized-controlled-trials-to-drive-effective-social-spending/.
- Leigh, Andrew. "Randomised Policy Trials." *Agenda* 10, no. 4 (2003): 341–54.
- Liebman, Jeffrey, and Alina Sellman. "Social Impact Bonds, A Guide for State and Local Governments." Harvard Kennedy School, Social Impact Bond Technical Assistance Lab, June 2013.
- Llewellyn, Gwynnyth, David McConnell, Anne Honey, Rachel Mayes, and Domenica Russo. "Promoting Health and Home Safety for Children of Parents with Intellectual Disability: A Randomized Controlled Trial." *Research in Developmental Disabilities* 24, no. 6 (November 2003): 405–31. doi:10.1016/j.ridd.2003.06.001.
- Manzi, Jim. *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*. New York: Basic Books, 2012.
- Markie-Dadds, Carol, and Matthew R. Sanders. "Self-Directed Triple P (Positive Parenting Program) for Mothers with Children at-Risk of Developing Conduct Problems." *Behavioural and Cognitive Psychotherapy* 34, no. 03 (July 2006): 259. doi:10.1017/S1352465806002797.
- Marsh, Herbert W., and Naida D. Peart. "Competitive and Cooperative Physical Fitness Training Programs for Girls: Effects on Physical Fitness and Multidimensional Self-Concepts." *Journal of Sport & Exercise Psychology* 10, no. 4 (1988): 390–407.
- Martin, Alicia J., and Matthew R. Sanders. "Balancing Work and Family: A Controlled Evaluation of the Triple P-Positive Parenting Program as a Work-Site Intervention." *Child and Adolescent Mental Health* 8, no. 4 (2003): 161–69. doi:10.1111/1475-3588.00066.
- Mazerolle, Lorraine, Lyndel Bates, Sarah Bennett, Gentry White, Jason Ferris, and Emma Antrobus. "Optimising the Length of Random Breath Tests: Results from the Queensland Community Engagement Trial." *Australian & New Zealand Journal of Criminology* 48, no. 2 (June 1, 2015): 256–76. doi:10.1177/0004865814532661.
- "Measure for Measure." *The Economist*, December 12, 2015. <http://www.economist.com/news/international/21679811-how-test-everything-sluggish-teenagers-corrupt-bureaucrats-measure-measure>.
- "Meta-Cognition and Self-Regulation." *Education Endowment Foundation*. Accessed March 1, 2016. <https://educationendowmentfoundation.org.uk/evidence/teaching-learning-toolkit/meta-cognition-and-self-regulation/>.
- Milne, Janet, and Susan H. Spence. "Training Social Perception Skills with Primary School Children: A Cautionary Note." *Behavioural Psychotherapy* 15, no. 2 (1987): 144–57. doi:10.1017/S014134730001123X.
- "Mind, Society, and Behavior - World Development Report." World Bank Group, 2015.
- Mitchell, Jo, Sandra Palmer, Michael Booth, and Gawaine Powell Davies. "A Randomised Trial of an Intervention to Develop Health Promoting Schools in Australia: The South Western Sydney Study." *Australian and New Zealand Journal of Public Health* 24, no. 3 (June 2000): 242–46.
- Morawska, Alina, and Matthew Sanders. "An Evaluation of a Behavioural Parenting Intervention for Parents of Gifted Children." *Behaviour Research and Therapy* 47, no. 6 (2009): 463–70. doi:10.1016/j.brat.2009.02.008.
- Morawska, Alina, and Matthew R. Sanders. "Self-Administered Behavioral Family Intervention for Parents of Toddlers: Part I. Efficacy." *Journal of Consulting and Clinical Psychology* 74, no. 1 (2006): 10–19. doi:10.1037/0022-006X.74.1.10.
- Mullainathan, Sendhil, and Michael Faye. "Field Experiments in Rural Finance: An Example from Tamil Nadu, India," March 21, 2007.
- Nagel, Tricia, Gary Robinson, John Condon, and Tom Trauer. "Approach to Treatment of Mental Illness and Substance Dependence in Remote Indigenous Communities: Results of a Mixed Methods Study." *Australian Journal of Rural Health* 17, no. 4 (August 1, 2009): 174–82. doi:10.1111/j.1440-1584.2009.01060.x.
- Burwell, Sylvia, Cecilia Munoz, John Holdren, and Alan Krueger. "Next Steps on the Evidence and Innovation Agenda," July 26, 2013.
- "NSW FACS Behavioural Insights Toolkit." NSW Behavioural Insights Unit, April 27, 2015.
- Olds DL, Holmberg JR, Donelan-McCall N, Luckey DW, Knudtson MD, and Robinson J. "Effects of Home Visits by Paraprofessionals and by Nurses on Children: Follow-up of a Randomized Trial at Ages 6 and 9 Years." *JAMA Pediatrics* 168, no. 2 (February 1, 2014): 114–21. doi:10.1001/jamapediatrics.2013.3817.
- Parker, Eleanor J, Gary Misan, Alwin Chong, Helen Mills, Kaye Roberts-Thomson, Alice M Horowitz, and Lisa M Jamieson. "An Oral Health Literacy Intervention for Indigenous Adults in a Rural Setting in Australia." *BMC Public Health* 12, no. 1 (2012): 461. doi:10.1186/1471-2458-12-461.
- Peel, Nancye, Margaret Steinberg, and Gail Williams. "Home Safety Assessment in the Prevention of Falls among Older People." *Australian and New Zealand Journal of Public Health* 24, no. 5 (October 2000): 536–39.
- Philliber S, Kaye JW, Herrling S, and West E. "Preventing Pregnancy and Improving Health Care Access among Teenagers: An Evaluation of the Children's Aid Society-Carrera Program." *Perspectives on Sexual & Reproductive Health* 34, no. 5 (October 9, 2002): 244–51 8p.

- Purcell, Amanda, Jennifer Fleming, Bryan Burmeister, Sally Bennett, and Terry Haines. "Is Education an Effective Management Strategy for Reducing Cancer-Related Fatigue?" *Supportive Care in Cancer* 19, no. 9 (2011): 1429–39. doi:10.1007/s00520-010-0970-2.
- "Randomised Control Trials: Coming of Age." *The Economist*, Accessed February 9, 2016. <http://www.economist.com/blogs/freeexchange/2013/12/randomised-control-trials>.
- "Request for Proposals: Low-Cost Randomized Controlled Trials to Drive Effective Social Spending." Laura and John Arnold Foundation, December 2015.
- Richardson, Ann, Sheila Williams, Mark Elwood, Margaret Bahr, and Tim Medlicott. "Participation in Breast Cancer Screening: Randomised Controlled Trials of Doctors' Letters and of Telephone Reminders." *Australian Journal of Public Health* 18, no. 3 (September 1994): 290.
- "Rigorous Program Evaluations on a Budget: How Low-Cost Randomized Controlled Trials Are Possible in Many Areas of Social Policy." Coalition for Evidence-Based Policy, March 2012.
- Rogers, Patricia, Andrew Hawkins, Bron McDonald, Alice Macfarlan, and Chris Milne. "Choosing Appropriate Designs and Methods for Impact Evaluation." Office of the Chief Economist, Department of Industry, Innovation and Science, Australian Government, November 2015.
- Rogers, Todd. "Low Cost/ High ROI Experiments in Organizations." *Unpublished*, 2016.
- Sanders, Matthew R., Carol Markie-Dadds, Lucy A. Tully, and William Bor. "The Triple P-Positive Parenting Program: A Comparison of Enhanced, Standard, and Self-Directed Behavioral Family Intervention for Parents of Children with Early Onset Conduct Problems." *Journal of Consulting and Clinical Psychology* 68, no. 4 (August 2000): 624–40. doi:10.1037/0022-006X.68.4.624.
- Sanders, Matthew R., Danielle T. Montgomery, and Margaret L. Brechman-Toussaint. "The Mass Media and the Prevention of Child Behavior Problems: The Evaluation of a Television Series to Promote Positive Outcomes for Parents and Their Children." *Journal of Child Psychology and Psychiatry* 41, no. 7 (October 1, 2000): 939–48. doi:10.1111/1469-7610.00681.
- Schofield, Margot J., Rob W. Sanson-Fisher, and Sally Gulliver. "Interventions with Retailers to Reduce Cigarette Sales to Minors: A Randomised Controlled Trial." *Australian and New Zealand Journal of Public Health* 21, no. 6 (October 1997): 590–96.
- Schultz, Cynthia L., Noel C. Schultz, Elizabeth J. Bruce, Kosmas X. Smyrniotis, Lindsay B. Carey, and Christa L. Carey. "Psychoeducational Support for Parents of Children with Intellectual Disability: An Outcome Study." *International Journal of Disability, Development and Education* 40, no. 3 (January 1, 1993): 205–16. doi:10.1080/0156655930400307.
- "Search Hansard – Parliament of Australia." Accessed March 1, 2016. [http://www.aph.gov.au/Parliamentary\\_Business/Hansard/Search](http://www.aph.gov.au/Parliamentary_Business/Hansard/Search).
- Shanahan, Marian, Emily Lancsar, Marion Haas, Bronwyn Lind, Don Weatherburn, and Shuling Chen. "Cost-Effectiveness Analysis of the New South Wales Adult Drug Court Program." *Evaluation Review* 28, no. 1 (February 1, 2004): 3–27. doi:10.1177/0193841X03257531.
- Shergold, Peter. "Learning from Failure." Commonwealth of Australia, August 12, 2015.
- Shu, Lisa L., Nina Mazar, Francesca Gino, Dan Ariely, and Max H. Bazerman. "Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End." *Proceedings of the National Academy of Sciences* 109, no. 38 (September 18, 2012): 15197–200. doi:10.1073/pnas.1209746109.
- Skouteris, Helen, Marita McCabe, Boyd Swinburn, and Briony Hill. "Healthy Eating and Obesity Prevention for Preschoolers: A Randomised Controlled Trial." *BMC Public Health* 10, no. 1 (2010): 220. doi:10.1186/1471-2458-10-220.
- Smith, Peter K., D. J. Pepler, and Ken Rigby. *Bullying in Schools : How Successful Can Interventions Be?* New York: Cambridge University Press, 2004.
- Smith, P. M., and E. Burgess. "Smoking Cessation Initiated during Hospital Stay for Patients with Coronary Artery Disease: A Randomized Controlled Trial." *Canadian Medical Association Journal* 180, no. 13 (June 23, 2009): 1297–1303. doi:10.1503/cmaj.080862.
- "Social and Behavioral Sciences Team Annual Report." Executive Office of the President National Science and Technology Council, September 2015.
- Sousa Lourenco, Joana, Emanuele Ciriolo, Sara Rafael Almeida, and Xavier Troussard. "Behavioural Insights Applied to Policy." European Commission, 2016. [http://publications.jrc.ec.europa.eu/repository/bitstream/JRC100146/kjna27726enn\\_new.pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC100146/kjna27726enn_new.pdf).
- "So You Want to Do an RCT with a Government\_ Things You Should Know — Running Randomized Evaluations\_ A Practical Guide.pdf," n.d.
- "Spotlight on Health Results Behavioural Insights Short Report." Premier & Cabinet, Behavioural Insights Unit, NSW, 2015.

- Strang, Heather, Lawrence Sherman, Daniel Woods, and Geoffrey Barnes. "Australian Institute of Criminology - Experiments in Restorative Policing: Final Report," November 2011. [http://www.aic.gov.au/criminal\\_justice\\_system/rjustice/rise/final.html](http://www.aic.gov.au/criminal_justice_system/rjustice/rise/final.html).
- Taft, Angela J, Rhonda Small, Kelsey L Hegarty, Lyndsey F Watson, Lisa Gold, and Judith A Lumley. "Mothers' AdvocateS In the Community (MOSAIC)- Non-Professional Mentor Support to Reduce Intimate Partner Violence and Depression in Mothers: A Cluster Randomised Trial in Primary Care." *BMC Public Health* 11, no. 1 (2011): 178. doi:10.1186/1471-2458-11-178.
- Thaler, Richard H. *Misbehaving: The Making of Behavioral Economics*. First edition. New York: W. W. Norton & Company, 2015.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Revised & Expanded edition. New York: Penguin Books, 2009.
- "The Behavioural Insights Team, Update Report 2013-2015." The Behavioural Insights Team, 2015.
- "The Campbell Collaboration." Accessed March 7, 2016. <http://www.campbellcollaboration.org/lib/>.
- The Campbell Collaboration, The Steering Group of. "Campbell Collaboration Systematic Reviews: Policies and Guidelines." *Campbell Systematic Reviews*, March 2, 2015. <http://www.campbellcollaboration.org/lib/project/328/>.
- "Top Tier Evidence." Accessed February 10, 2016. <http://toptierevidence.org/>.
- "Treatment Foster Care Oregon." *EPISCenter*. Accessed February 10, 2016. <http://www.episcenter.psu.edu/ebp/multidimensional>.
- "Turnbull Government Creates New 'Behavioural Economics' Team of Advisers." *The Sydney Morning Herald*. Accessed February 7, 2016. <http://www.smh.com.au/federal-politics/political-news/turnbull-government-creates-new-behavioural-economics-team-of-advisers-20151123-gl5j6w.html>.
- Turner, Karen Mt, Mary Richards, and Matthew R. Sanders. "Randomised Clinical Trial of a Group Parent Education Programme for Australian Indigenous Families." *Journal of Paediatrics and Child Health* 43, no. 6 (2007): 429–37. doi:10.1111/j.1440-1754.2007.01053.x.
- "Understanding People, Better Outcomes - Behavioural Insights in NSW." NSW Premier & Cabinet, 2015.
- "Victoria Gives Behavioural Insights a Nudge, Cross-Agency Mandate." *The Mandarin*, March 23, 2016. <http://www.themandarin.com.au/62107-victoria-gives-behavioural-insights-unit-nudge/>.
- Waters, Elizabeth, Rebecca Armstrong, Boyd Swinburn, Laurence Moore, Maureen Dobbins, Laurie Anderson, Mark Petticrew, et al. "An Exploratory Cluster Randomised Controlled Trial of Knowledge Translation Strategies to Support Evidence-Informed Decision-Making in Local Governments (The KT4LG Study)." *BMC Public Health* 11 (2011): 34. doi:10.1186/1471-2458-11-34.
- Wen, Li Ming, Louise A. Baur, Judy M. Simpson, Chris Rissel, and Victoria M. Flood. "Effectiveness of an Early Intervention on Infant Feeding Practices and 'Tummy Time': A Randomized Controlled Trial." *Archives of Pediatrics & Adolescent Medicine* 165, no. 8 (August 2011): 701–7. doi:10.1001/archpediatrics.2011.115.
- Wen, Li Ming, Louise A. Baur, Judy M. Simpson, Chris Rissel, Karen Wardle, and Victoria M. Flood. "Effectiveness of Home Based Early Intervention on Children's BMI at Age 2: Randomised Controlled Trial." *BMJ* 344 (June 26, 2012): e3732. doi:10.1136/bmj.e3732.
- Wen, Li Ming, Louise A. Baur, Judy M. Simpson, Huilan Xu, Alison J. Hayes, Louise L. Hardy, Mandy Williams, and Chris Rissel. "Sustainability of Effects of an Early Childhood Obesity Prevention Trial Over Time: A Further 3-Year Follow-up of the Healthy Beginnings Trial." *JAMA Pediatrics* 169, no. 6 (June 2015): 543–51. doi:10.1001/jamapediatrics.2015.0258.
- Wen, Li Ming, Denise Fry, Dafna Merom, Chris Rissel, Helen Dirakis, and Angela Balafas. "Increasing Active Travel to School: Are We on the Right Track? A Cluster Randomised Controlled Trial from Sydney, Australia." *Preventive Medicine* 47, no. 6 (December 2008): 612–18. doi:10.1016/j.ypmed.2008.09.002.
- "Who We Are" *The Behavioural Insights Team*, Accessed February 7, 2016. <http://www.behaviouralinsights.co.uk/about-us/>.
- "Winning Entry: Becoming a Man (B.A.M.)." *Crime Lab, The University of Chicago*. Accessed March 28, 2016. <https://crimelab.uchicago.edu/page/Winning-Entry-BAM-Sports-Edition>.